

# UNIVERSIDAD CARLOS III DE MADRID

ESCUELA POLITÉCNICA SUPERIOR

DEPARTAMENTO DE TEORÍA DE LA SEÑAL Y COMUNICACIONES



PROYECTO FIN DE CARRERA

## DISEÑO E IMPLEMENTACIÓN DE UN DETECTOR AUTOMÁTICO DE ACORDES

INGENIERÍA TÉCNICA DE TELECOMUNICACIONES: SONIDO E IMAGEN

Autor: **Julio Martín Macías**

Tutores: **Emilio Parrado Hernández**

**Raúl Santos Rodríguez**

Leganés (Madrid), Junio de 2012



## **Resumen**

En este trabajo se diseña e implementa un sistema de reconocimiento automático de acordes a partir de una muestra de audio musical.

El diseño del reconocedor se basa en un **modelo estadístico de detección analítica**.

La primera parte se dedica a la creación de una base de datos de música etiquetada compuesta por muestras grabadas de acordes, tocados en distintos instrumentos musicales, y canciones de *Michael Jackson* y *Beatles*.

Además, se detalla cómo se ha llevado a cabo la grabación de las muestras de instrumentos y el proceso de etiquetado de canciones.

Posteriormente se desarrolla el detector de acordes mediante un estimador de máxima verosimilitud implementado en el software matemático *Matlab*.

Dicha detección pasará por un bloque de postprocesado que suavizará los cambios bruscos entre acordes basándose en la capacidad cognitiva humana.

Los resultados de la evaluación de la detección muestran un comportamiento satisfactorio en el reconocimiento de acordes en las muestras grabadas. En el reconocimiento de acordes en canciones se produce una contribución positiva, pero es necesaria una mejora sustancial para obtener resultados más fiables. Sin embargo, se logran identificar patrones que se producen cuando la detección es errónea, tales como la confusión entre el modo mayor y menor o la confusión entre acordes que son relativos entre sí. Además se demuestra la disminución de la tasa de error en la detección al incluir un bloque de postprocesado en forma de filtro de modas.



### **Agradecimientos.**

*A mi familia, por el apoyo constante, a mis compañeros de clase, los “jambooooos”, y a cada uno de los profesores que han contribuido a que haya llegado a escribir este trabajo, en especial a Emilio y a Raúl por su gran ayuda.*



## Índice

<b>Capítulo 1. Introducción</b>	11
1.1 Motivación	12
1.2 Objetivos	13
1.3 Organización del documento	13
<b>Capítulo 2. Revisión del estado del arte</b>	17
2.1 Conceptos musicales y terminología	17
2.1.1 Notas musicales	17
2.1.2 Acordes	19
2.1.3 Terminología musical	21
2.1.4 Acordes relativos	22
2.1.5 Utilización de acordes en la música	22
2.2 Sobre el reconocimiento de acordes	24
2.2.1 Bajo nivel	24
2.2.1.1 Análisis de audio	25
2.2.1.2 Modelos de acordes en trabajos previos	29
2.2.2 Alto nivel	30
2.2.2.1 Detección analítica	30
2.2.2.2 Detección de acordes en trabajos previos	33
2.2.3 Métodos de evaluación de la detección	33
2.3 Aplicaciones comerciales	34
<b>Capítulo 3. Diseño e implementación</b>	39
3.1 Introducción	39
3.2 Creación de la base de datos	40
3.2.1 Elección de las muestras de audio	40
3.2.1.1 Acordes aislados	41
3.2.1.2 Canciones	42
3.2.2 Análisis frecuencial de las muestras	44
3.2.2.1 Analizador. NNLS Chroma	44
3.2.3 Etiquetado de las muestras	45

3.3	Prototipo .....	49
3.3.1	Introducción a la base de datos .....	49
3.3.1.1	Carga de matrices Song .....	49
3.3.1.2	Carga de matrices Chord .....	49
3.3.1.3	Alineamiento de las matrices de etiquetado .....	51
3.3.2	Extracción de datos de entrenamiento .....	53
3.3.2.1	Agrupación de datos de entrenamiento .....	53
3.3.2.2	Cálculo de medias y varianzas .....	53
3.3.3	Detector .....	54
3.3.3.1	Detector ML .....	54
3.3.3.2	Postprocesado .....	55
3.3.4	Evaluación de la detección .....	57
3.3.4.1	Probabilidad de error .....	57
3.3.4.2	Matriz de confusión .....	57
<b>Capítulo 4. Valoración experimental .....</b>		<b>59</b>
4.1	Cromograma de muestras .....	60
4.1.1	Detección de acordes aislados .....	63
4.1.1.1	Entrenamiento con muestras de piano Rhodes .....	63
4.1.1.2	Entrenamiento con muestras de piano acústico .....	64
4.1.1.3	Entrenamiento con muestras de guitarra .....	64
4.1.1.4	Entrenamiento con muestras de los tres instrumentos .....	65
4.1.2	Detección de acordes en canciones .....	66
4.1.2.1	Entrenamiento con acordes aislados .....	67
4.1.2.2	Entrenamiento con acordes aislados y Beatles .....	67
4.1.2.3	Entrenamiento con acordes aislados y Michael J .....	68
4.1.2.4	Entrenamiento con Beatles .....	68
4.1.2.5	Entrenamiento con Michael J .....	69
4.1.2.6	Entrenamiento con Michael J. y Beatles .....	69
4.1.2.7	Entrenamiento con acordes aislados, Michael J. y Beatles....	70
4.1.2.8	Observaciones .....	70
4.2	Influencia del porcentaje del conjunto de entrenamiento .....	73
4.3	Influencia del postprocesado .....	74
4.3.1	Tamaño de ventana .....	75



4.4	Matrices de confusión .....	77
4.4.1	Acordes aislados .....	78
4.4.1.1	Entrenamiento con piano Rhodes .....	78
4.4.1.2	Entrenamiento con guitarra .....	79
4.4.1.3	Entrenamiento con piano acústico .....	80
4.4.2	Acordes en canciones .....	81
4.4.2.1	Entrenamiento con acordes aislados, Michael J. y Beatles ....	81
4.4.2.2	Entrenamiento con Michael J .....	83
4.4.2.3	Entrenamiento con Beatles .....	84
4.5	Valoración de las pruebas .....	85
	<b>Capítulo 5. Conclusiones</b> .....	<b>87</b>
5.1	Trabajo futuro.....	88
	<b>Referencias</b> .....	<b>91</b>
	<b>Presupuesto</b> .....	<b>95</b>



## **Capítulo 1. Introducción**

El reconocimiento automático de características musicales a partir de una onda sonora es un campo en desarrollo constante. La principal distinción frente a detecciones en otros campos, como el de la imagen, es que toda la información se concentra en una sola onda longitudinal. A pesar de ello, la cantidad de información que puede extraerse es inmensa. Además de las cuestiones músico-teóricas, se pueden implementar detectores del estilo musical, ritmo e incluso de sensaciones provocadas en el ser humano. Incluso toda esa información puede complementarse entre sí para el desarrollo de transcriptores musicales, buscadores de canciones, etc.

La parte más objetiva, dentro de todo este entramado de características, es la que engloba la teoría musical. En ella podemos encontrar conceptos como melodía, tempo o acompañamiento.

La **melodía** se refiere a la sucesión de notas musicales preponderante en un fragmento musical, que se desarrolla generalmente en la parte media-aguda del ancho de frecuencia del tema.

El **tempo** marca la velocidad con la que debe ejecutarse una pieza musical.

Estas características son relativamente fáciles de reconocer por cualquier persona aficionada a la música. Todos hemos cantado o silbado una melodía sin conocer con exactitud qué notas musicales se estaban ejecutando, o hemos dado palmas al “ritmo de la música” sin conocer la velocidad en términos de unidades de tempo.

Sin embargo, el **acompañamiento musical**, que se sustenta sobre el tempo y acompaña a la melodía, es más difícil de reconocer por un ser humano, aun poseyendo un mínimo de conocimiento teórico-musical.

El acompañamiento musical es un concepto subjetivo que engloba muchas características. No obstante, hay elementos primarios que, en conjunto, rigen la forma que debe seguir. Estos elementos son los **acordes**.

El reconocimiento de acordes en una pieza musical es una tarea que requiere el entrenamiento auditivo y la educación musical del individuo, por lo que es necesario un experto musical para llevarlo a cabo.

Precisamente por ese motivo, los reconocedores de acordes y de formas musicales desarrollados hasta ahora intentan simular de forma automática el proceso que sigue un individuo al identificar acordes.

Para ello se utiliza un modelo estadístico de detección analítica. Este método engloba una serie de pasos hasta llegar a la detección automática de acordes.

Primero se elige la base de datos compuesta de muestras de audio musicales con la que entrenar el prototipo. Una vez entrenado se implementa el detector y se hacen las primeras detecciones de acordes.

Por último, se evalúa la calidad de la detección con una serie de herramientas. En función de los resultados, se pueden producir modificaciones tanto la etapa de entrenamiento como en el desarrollo del detector con el fin de mejorar la calidad en el reconocimiento.

Este proyecto se centra tanto en el desarrollo de la base de datos, como en el diseño e implementación de un prototipo capaz de reconocer acordes en muestras de audio musicales. Además, se presentan los resultados de la detección en función de diferentes parámetros.

## 1.1 Motivación.

El reconocimiento de acordes automático está orientado a agilizar el proceso a un músico inexperto que desee obtener la sucesión de acordes de una canción de forma casi inmediata.

Hoy en día, se ha multiplicado el acceso a repositorios de música digitalizada y es posible el manejo de grandes cantidades de información para el desarrollo de bases de datos con las que entrenar y testear un reconocedor automático.

Además, un reconocedor en tiempo real puede ayudar en labores pedagógicas con el fin de reconocer fallos de forma inmediata en los acordes en un aprendizaje de instrumento.

Por último, se puede complementar con un detector de melodía o de tempo con el fin de realizar una transcripción más completa de una canción.

## 1.2 Objetivos

El objetivo de este proyecto es diseñar e implementar los pasos necesarios para el desarrollo de un reconocedor de acordes.

Estos pasos se dividen en:

- Creación de una base de datos que se componga de muestras de audio etiquetadas con los acordes correspondientes. Para ello, se utilizará una base de datos existente de canciones del grupo *Beatles*. Por otra parte, se ampliará con el etiquetado de canciones del artista *Michael Jackson* y se grabarán muestras de instrumentos ejecutando acordes aislados con el fin de estudiar su influencia en el reconocimiento.
- Implementación de un modelo estadístico de detección en el software matemático *Matlab*. La detección de acordes se llevará a cabo basándose en el modelo de estimación de máxima verosimilitud.
- Desarrollo de un bloque de postprocesado con el fin de filtrar posibles detecciones erróneas.
- Evaluación de los resultados en base al porcentaje de error de detección y la matriz de confusión. Con estas herramientas se estudiará la influencia de la elección de distintos conjuntos de entrenamiento y la inclusión del bloque de postprocesado.

## 1.3 Organización del documento.

### Capítulo 2. Revisión del estado del arte.

Este capítulo está dividido en tres apartados:

En el apartado 2.1 se explican los conceptos musicales básicos para facilitar el entendimiento del texto. En primer lugar, se definen las notas musicales desde el punto de vista físico-teórico. A continuación, se definen los acordes, se indica la terminología usada y se da una breve explicación del papel que desempeñan en la música.

En el apartado 2.2 se desglosan los pasos que se deben seguir en el desarrollo de un reconocedor.

- El desarrollo de *bajo nivel* (apartado 2.2.1) detalla las formas que han adoptado diferentes autores en trasladar la definición musical de los acordes a términos físicos, con el objetivo de hacer más eficiente su reconocimiento.
- En el desarrollo de *alto nivel* (apartado 2.2.2) se detallan los fundamentos matemáticos para el desarrollo de un reconocedor de acordes basado en la detección analítica y las formas de detección adoptadas por otros autores.
- En el apartado 2.2.3 se especifican las herramientas utilizadas para evaluar la calidad de la detección y se define la Matriz de confusión como una de ellas.

Por último, en el apartado 2.3 se muestran diferentes ejemplos de aplicaciones comerciales relacionadas con el reconocimiento de formas musicales en el audio. Entre ellas se encuentra un detector de melodía, un afinador polifónico y ejemplos de detectores de acordes.

### Capítulo 3. Diseño e implementación.

En este capítulo se enumeran y se detallan los pasos que se han seguido en el desarrollo del reconocedor de acordes, modelado como un problema de detección analítica.

- En el apartado 3.1 se hace una descripción en líneas generales del desarrollo del prototipo, descrito detalladamente en los siguientes apartados.
- En el apartado 3.2 se justifica la elección de la base de datos sobre la que entrenar y evaluar el reconocedor. Se detalla tanto el proceso de grabación de determinadas muestras de audio como el etiquetado de todas ellas.
- Una vez dispuesta la base de datos, en el apartado 3.3 se especifica el desarrollo del detector en forma de estimador de máxima verosimilitud y de elementos que evalúan su calidad. Todo ello desarrollado en el software matemático *Matlab*.

### Capítulo 4. Valoración experimental.

En este capítulo se presenta el resultado de las pruebas de evaluación de la detección de acordes sobre muestras de audio de la base de datos. También se hacen pruebas para determinar la influencia de algunas variables que intervienen en éste, con el fin de calcular su valor óptimo para minimizar el error.

Primero, se calcula el porcentaje de error en estimaciones de muestras con diferentes conjuntos de entrenamiento. A continuación, se analizan las matrices de confusión sobre estas estimaciones con el objetivo de identificar patrones en la detección errónea de acordes.

## Capítulo 5. Conclusiones

Por último, se resumen las conclusiones en base a los resultados que se han obtenido en la elaboración de este proyecto.

Además se indican posibles mejoras para el prototipo y otras vías sobre las que continuar el proyecto, como el desarrollo de una interfaz gráfica o la inclusión de éste en un detector musical más sofisticado.





## Capítulo 2. Revisión del estado del arte

Un conocimiento básico de la teoría musical es un requisito para el desarrollo y el entendimiento del prototipo que se describe a continuación. Este capítulo comienza con una breve explicación de los conceptos musicales básicos. Además, se desglosan los pasos que normalmente se siguen en el desarrollo de un reconocedor de acordes y por último, se aporta algún ejemplo comercial utilizado en la actualidad.

### 2.1 Conceptos musicales y terminología.

#### 2.1.1 Notas Musicales

En la rama de la física del sonido, una nota musical se refiere a un sonido determinado por una vibración cuya frecuencia fundamental es constante.

En el sistema de música occidental se ha acordado tomar sólo unas frecuencias concretas. Con tales frecuencias se definen 12 notas musicales.

Nota	Frecuencia (Hz)
Do	261,63
Do#/Reb	277,18
Re	293,66
Re#/Mib	311,13
Mi	329,63
Fa	349,23
Fa#/Solb	369,99
Sol	392
Sol#/Lab	415,30
La	440
La#/Sib	466,16
Si	493,88

- *b*: Bemol.
- *#*: Sostenido.

*Tabla 1. Frecuencia fundamental de las notas.*

La frecuencia fundamental de las notas es aproximadamente proporcional en un eje logarítmico-frecuencial.

La distancia entre dos notas musicales se denomina intervalo, y se mide en “semitonos”. Un **semitono** es la distancia que existe entre dos notas consecutivas. Se habla de **intervalo ascendente** al intervalo orientado hacia frecuencias superiores. Análogamente, se habla de **intervalo descendente** cuando se produce hacia frecuencias inferiores.

Esas 12 notas musicales se dividen en:

- a) **Notas diatónicas (7)**: Do, Re, Mi, Fa, Sol, La y Si.
- b) **Notas alteradas (5)**: Se les llama alteradas porque representan una alteración de las notas diatónicas. Una alteración se refiere al hecho de aumentar o disminuir un semitono a una nota diatónica. Para aumentar se utiliza el símbolo **#** (sostenido) y para disminuir el símbolo **b** (bemol) acompañando al nombre de la nota diatónica a la que alteran. Existen 5 notas alteradas y cada una de ellas puede nombrarse de dos maneras. Do#/Reb, Re#/Mib, Fa#/Solb, Sol#/Lab y La#/Sib.

En la Figura 1 de la siguiente página se da una explicación gráfica a este apartado.

La sucesión de notas se repite periódicamente cada 12 semitonos. Al conjunto de 12 semitonos se le denomina octava. La frecuencia de cada nota en función de la octava en la que se encuentra se puede calcular como:

$$F_{nota} = F_{ref} * 2^n$$

Siendo:

- **Fref** corresponde a la frecuencia de referencia, presentada en la Tabla 1.
- **n** representa el número de la octava y suele tomar valores enteros en el intervalo de (-3,3). Valores inferiores y superiores suponen notas demasiado graves y agudas respectivamente para su utilización en la música. Cuando n=0 se habla de “octava central”.

Para ilustrar esta explicación se utilizará como ejemplo la distribución de las notas en un piano (Figura 1).

En éste, las notas diatónicas son de color blanco y las notas alteradas de color negro. La abreviatura “st” se refiere a la distancia de un semitono.

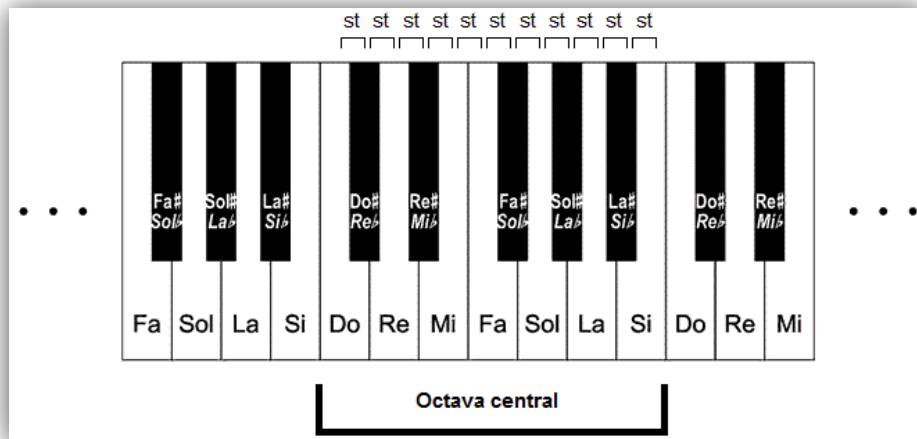


Figura 1. Notas musicales en un piano

Teóricamente, el espectro audible por el ser humano es de 20Hz a 20 KHz. La gama de frecuencias fundamentales con las que se trabaja en la música es de 27 a 13.200 Hz (13.186 Hz es la frecuencia fundamental de la última nota del piano). La región superior (hasta los 20 KHz) está dedicada a los armónicos de los tonos altos, que dan el timbre de los diferentes instrumentos. Más información en: [Recuero 1999]

### 2.1.2 Acordes

En teoría musical, un **acorde** consiste en un determinado conjunto de notas diferentes que suenan **simultáneamente o en sucesión rápida**, y que constituyen una identidad o unidad armónica. Las notas pueden pertenecer a la misma o a diferentes octavas.

Está definido por dos términos:

- Nota fundamental o tónica:** Determina la nota principal del acorde.
- Modo:** Define, a partir de la nota fundamental, las notas restantes que forman el acorde. Estas notas van a encontrarse a una determinada distancia (o intervalo musical) de la nota fundamental. Cada modo aporta una sonoridad característica al acorde. Alegría, melancolía, tristeza, tensión, ...

Las combinaciones de dos notas se llaman diadas, las de tres, triadas, las de cuatro, cuatriadas, las de cinco, quintiadas, las de seis, sextiadas, y las de siete, septiadas.

Pueden hacerse todo tipo de combinaciones con las notas que existen, pero no todas se consideran acordes (Aunque, a medida que la música ha evolucionado, se han ido considerando nuevos).

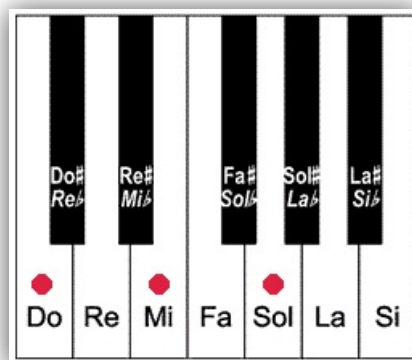
Teóricamente, la **combinación mínima** de notas que se necesita para crear un acorde es de **tres**.

Para las **triadas** existen hasta seis modos diferentes, pero los más representativos y los que se utilizarán en el desarrollo del proyecto son dos:

a) **Modo mayor:**

A partir de una tónica, las dos notas restantes que forman el acorde están a una distancia de 4 y 7 semitonos.

Por ejemplo, un acorde con la tónica en Do y de modo mayor (Figura 2) está formado por la tónica (Do), una nota a distancia de 4 semitonos (Mi) y otra a distancia de 7 semitonos (Sol).



*Figura 2. Acorde de Do Mayor*

b) **Modo menor:**

A partir una tónica dada, las dos notas restantes que forman el acorde están a una distancia de 3 y 7 semitonos.

Un acorde con la tónica en Do y de modo menor (Figura 3) está formado por la tónica (Do), una nota a distancia de 3 semitonos (Mib) y otra a distancia de 7 semitonos (Sol).

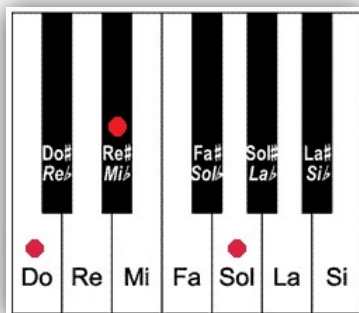


Figura 3. Acorde de Do menor

### 2.1.3 Terminología musical.

En este apartado se **describe** la terminología músico-teórica comúnmente usada. [Harte and Sander 2005]

#### a) Cifrado de notas:

El cifrado de notas que se suele adoptar cuando se trata de acordes es el anglosajón.

Tradicional	Cifrado anglosajón
La	A
Si	B
Do	C
Re	D
Mi	E
Fa	F
Sol	G

Tabla 2. Cifrado anglosajón de notas.

#### b) Cifrado de acordes:

Un acorde viene dado por la tónica y su modo. Se nombra como:

“Acorde de **Tónica Modo**”

De tal forma que un **Acorde de Do mayor** denota un acorde con la tónica en Do y de modo mayor.

En cuanto a la forma de escribirlo, se especifica en primer lugar la tónica en cifrado anglosajón, y le sigue una extensión determinada dependiendo de su modo.

En este prototipo, la extensión *maj* corresponde al modo mayor y *min* al modo menor.

De tal forma, un acorde de Do mayor se escribirá como **Cmaj** y un acorde de Do menor como **Cmin**.

#### 2.1.4 Acordes relativos.

Los acordes relativos, en triadas, se refieren a determinadas parejas de acordes en las que coinciden dos notas de las tres que lo forman. Además, las notas en las que difieren no crean disonancia<sup>1</sup> entre sí. Esta característica hace que puedan tocarse ambos acordes simultáneamente y dar lugar a confusión en su reconocimiento.

Las triadas de modo menor tienen un relativo de modo mayor a un intervalo (entre las tónicas de los acordes) de 3 semitonos de forma ascendente. Esta asociación es recíproca (en el apartado 2.1.1 se ha definido el concepto de intervalo).

Por ejemplo:

*Amin* (La Do Mi) tiene como relativo el acorde de *Cmaj* (Do Mi Sol), al igual que *Cmaj* tiene como relativo el acorde de *Amin*.

#### 2.1.5 Utilización de acordes en la música.

En las canciones o en temas instrumentales se pueden distinguir dos partes que suenan simultáneamente: La **melodía**, generada por la voz o un determinado instrumento solista, y la **música de acompañamiento**.

En un tema instrumental o cantado, las sucesiones de acordes, también llamadas **progresiones armónicas**, son el esquema que rige la forma que debe seguir la música de acompañamiento.

---

<sup>1</sup> Disonancia musical: Cualidad de aquellos sonidos que, oídos simultáneamente, producen efecto desagradable.

“Además, constituyen normalmente un **elemento invariable** entre diferentes versiones o variaciones, con los mismos o con diferentes instrumentos, de un mismo tema musical. Por ello, es típico “resumir” el acompañamiento de una canción en sus acordes básicos para luego poder interpretarla libremente al piano o la guitarra con dichos acordes como guía, sin perder la esencia de la composición, es decir, sin que esta deje de ser identificable. En sentido inverso, también es típico componer una canción para uno o varios instrumentos partiendo de la creación de una progresión de acordes”. [Wikipedia.org]

En el acompañamiento musical, los acordes pueden ser tocados por un instrumento o varios. En el caso de que los hagan sonar varios instrumentos, cada uno de ellos puede ejecutar diferentes notas, con la finalidad que el oyente capte la suma de todas ellas y perciba el acorde deseado.

Si se quiere más información sobre teoría musical, se puede consultar en: [Herrera 1984]

La Figura 4 muestra un ejemplo de partitura reducida. La sucesión de acordes (1) se escribe en la parte superior de las notas de la melodía que acompaña. (2). Se pueden encontrar más partituras de este formato en: [Johansson 1998]

**LADY MADONNA** - LENNON/MCCARTNEY

**VERSE 1**

A D A D

LADY WHO FINDS MA- DONNA CHILDREN AT YOUR FEET  
WHO FINDS THE MONEY WHEN YOU PAY THE RENT

**2**

A D E F G A

WONDER HOW YOU MANAGE TO MAKE ENDS MEET  
DID YOU THINK THAT MONEY WAS HER JEN- SENT

**BRIDGE**

Dm Dm/C Dm/B Dm/A G7 G7/F G7/E G7/D

FRIDAY NIGHT AR- RIVES WITHOUT A SUIT- CASE

C C/B C/A C/G Am Am/B Am/C Am/E

SUNDAY MORNING CREEPING LIKE A NUN

Dm /C /B /A G7 /F /E /D

MONDAY'S CHILD HAS LEARN'T TO TIE HIS SHOE- LACE

C BM7 E7sus E7

SEE HOW THEY RUN

VERSE: Lady Madonna, baby at your breast, Wonder how you manage to feed the rest?  
VERSE: instrumental  
BRIDGE: instrumental  
VERSE: Lady Madonna, lying on the bed, Listen to the music playing in your head  
VERSE: instrumental  
BRIDGE: Tuesday afternoon is never ending, Wednesday morning papers didn't come  
Thursday night your stockings needed mending, See how they run  
VERSE: Lady Madonna, children at your feet, Wonder how you manage to make ends meet?

Figura 4. Partitura de “Lady Madonna”. Beatles.

## 2.2 Sobre el reconocimiento de acordes.

En la elaboración de un reconocedor de acordes se pueden distinguir dos partes, descritas más detalladamente en: [Mauch 2010]

El desarrollo en **bajo nivel** se refiere a la forma adoptada en la descripción de los acordes.

El **alto nivel** agrupa todas las técnicas de detección de éstos en una muestra de audio musical.

### 2.2.1 Bajo nivel.

Un **modelo de acorde** determina la forma en la que se describe cada acorde. Suele ser en términos de parámetros armónicos de bajo nivel. Un ejemplo de este tipo de parámetros es el resultado de un análisis frecuencial de una muestra de audio.

En la **simbología musical**, basándose en la teoría, un acorde está expresado como combinación de eventos discretos de notas; es decir, en una triada se especifican únicamente las tres notas que lo forman. En el **análisis del audio**, sin embargo, se trabaja con características continuas de bajo nivel extraídas de la señal. La necesidad de adaptación de estas dos formas de expresión provoca diferencias entre los distintos enfoques que han dado los autores a los modelos de acordes. En los siguientes párrafos se analizará la evolución de las dos formas separadas.

#### a) **Simbología.**

Entre los primeros autores relacionados con éste tema, en el documento [Ulrich 1977] se representa un acorde como una lista de distancias en semitonos desde la raíz. Por ejemplo, Dmin sería D (3 7). El principal inconveniente al utilizar este método es que se limita a que las notas del acorde se encuentren en una misma octava. Y como se ha indicado previamente (apartado 2.1.2), las notas en un acorde pueden encontrarse en diferentes octavas.

Los trabajos posteriores de [Temperley and Sleator 1999], [Pardo 2002] y [Raphael and Stoddard 2003] (por mencionar algunos) solucionan el problema al aunar la información de todas las octavas en una sola, pero siguen representando el acorde únicamente como combinación de las notas que los componen. Los trabajos de [Fujishima 1999], [Wakefield 1999] y [Paiement et al. 2005] consistían en representar cada acorde como un vector de 12 componentes de números reales y no sólo por las



notas teóricas que contiene. Se observará en la explicación posterior, que ésta será la forma más adoptada a la hora de representar los acordes.

#### b) **Análisis del audio.**

Los modelos de acordes puramente teóricos, que se utilizan en la música escrita, no son fácilmente transferibles al dominio del audio.

Un simple espectro frecuencial no solo exhibe alta energía en la frecuencia fundamental, sino también en otras frecuencias relacionadas con ésta (P.ej. los armónicos superiores que se producen en la vibración de una cuerda). Además, en determinadas muestras de audio se pueden encontrar elementos que no hagan sonar ninguna nota del acorde en particular. Puede ser el caso de la voz e instrumentos que ejecutan la melodía del tema, instrumentos de percusión (los componentes de una batería) o efectos de sonido. Todos estos elementos pueden generar alta energía en frecuencias que no guardan relación con las notas del acorde.

Como resultado, los métodos de transcripción musical funcionan de forma fiable en determinados campos, como es la estimación de melodías monofónicas, aunque sigue siendo un tema en desarrollo.

Por tanto, es necesario encontrar la forma de expresar cada acorde en términos de parámetros extraídos de la señal de audio. Para [Fujishima 1999] cada acorde se puede modelar como un “pitch class profiles” (PCP), o también llamados “chroma vectors” (Vectores de cromograma) por [Wakefield 1999]. Cada PCP es una representación en 12 dimensiones del contenido armónico de una señal de audio. Cada dimensión representa la potencia de cada una de las doce notas musicales (independientemente de la octava en la que se encuentren). Para calcular tales potencias, el espectro se divide en bandas cuya frecuencia central coincide con las frecuencias de las notas musicales.

Esta representación en 12 dimensiones de la señal de audio se conoce como **cromograma**.

##### 2.2.1.1 Análisis del audio.

Para el cálculo del cromograma, es necesario obtener primero la información de su espectro. Esta información nos la proporciona el espectrograma.

a) **Cálculo del espectrograma.**

$X_k$  es el espectro de una señal de audio enventanada  $X[n]$ . Ese espectro se calcula usando la *transformada discreta de Fourier (DFT)*.

$$X_k = \sum_{n=0}^{N_F-1} x_n w_n e^{-\frac{2\pi i}{N_F} kn}, \quad k = 0, \dots, N_F - 1, \quad (1)$$

Donde  $N_F$  es el número de muestras en una ventana y  $w_n$  es una función ventana que pondera las muestras para reducir la filtración espectral asociada con los intervalos de observación finitos. [Oppenheim 1997]

b) **Cálculo del cromograma.**

La variable  $p$ , que toma valores de 0 a 11, se corresponde con cada una de las dimensiones, asociadas éstas a cada una de las 12 notas musicales.

El valor del **PCP** para cada  $p$  es calculado como el sumatorio de todas las potencias de los coeficientes del espectro cercanas a una instancia de esa clase o nota musical sin importar en qué octava se encuentre.

$$\text{PCP}_p = \sum_{M(m)=p} \|X_m\|^2, \quad (2)$$

Dónde:

$$M(m) = \text{round} \left( 12 \log_2 \left( \frac{f_s}{f_{\text{ref}}} \cdot \frac{m}{N_F} \right) \right) \bmod 12, \quad (3)$$

Con  $f_{ref}$  como frecuencia de referencia o frecuencia central de clase y  $f_s$  como frecuencia de muestreo. Usualmente, el espectro se calcula con ventanas con solapamiento. Este proceso también es conocido como *short-time Fourier transform* (STFT).

El espectro de frecuencia calculado por la Transformada de Fourier da como resultado una representación lineal del espectro, es decir, las frecuencias están equiespaciadas. Sin embargo, la relación de las frecuencias fundamentales de las notas musicales son proporcionales en un espacio frecuencial logarítmico (apartado 2.1.1). Debido a eso, es necesaria una conversión logarítmica en el eje frecuencial. Más información en: [Recuero 1999]

La matriz resultado  $(X_k, m)$ , en la cual la DFT de la ventana  $m_i$  ocupa la columna  $m_i$ , es el **espectrograma**. Al igual que el espectrograma describe el contenido espectral de una señal a lo largo del tiempo, la matriz del **cromograma**  $(PCP_p, m)$ , en el cual el vector de croma de la ventana  $m_i$  ocupa la columna  $m_i$ , describe el contenido cromático a lo largo del tiempo. A fin de permitir las diferencias de energía entre ventanas, cada vector del cromograma está normalizado.

A continuación se muestra el espectrograma (Figura 5) y el cromograma (Figura 6) de una misma canción. En la Figura 7 se muestra como ejemplo el cromograma de un acorde de Cmaj (Do mayor) ejecutado en un piano.

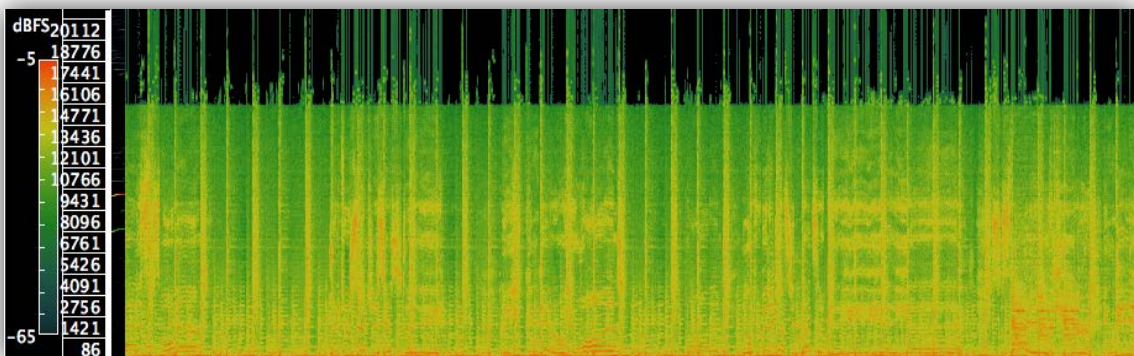
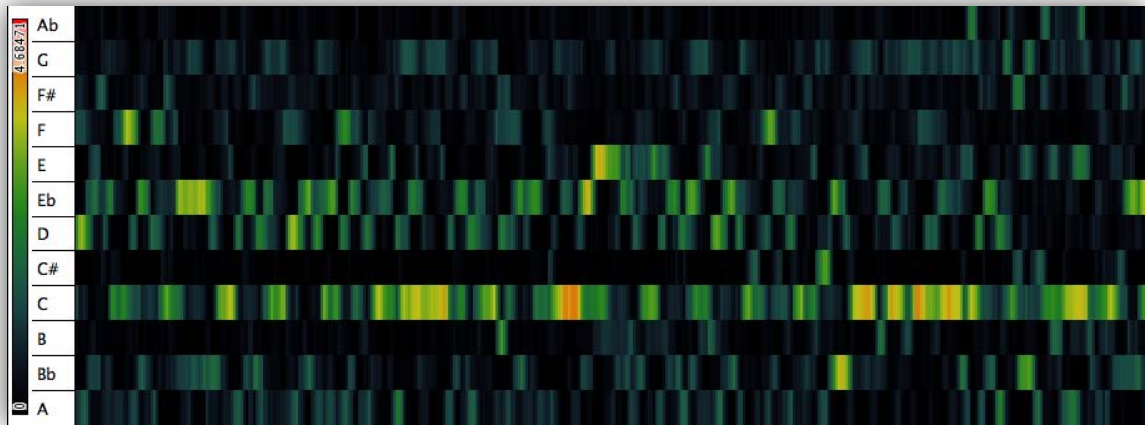
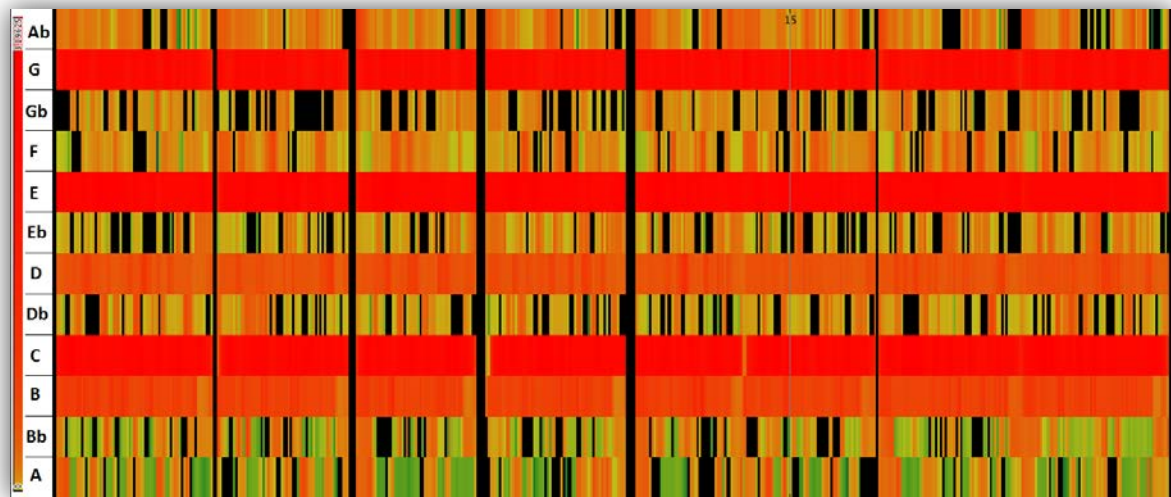


Figura 5. Espectrograma.



*Figura 6. Cromograma.*



*Figura 7. Cromograma de un Cmaj tocado en un piano.*

En el piano se han tocado las tres notas que forman el acorde de manera simultánea. C, E y G (Do, Mi y Sol).

Se puede observar que, además de las componentes principales del acorde, se producen excitaciones en otras frecuencias. En concreto, existe una excitación a nivel constante de otras notas que no se han tocado (D y B).

Según la teoría de aparición de armónicos en una cuerda, además de la frecuencia fundamental que produce la vibración de la cuerda, se excitarán otras frecuencias relacionadas con ésta.

En términos musicales, los dos primeros armónicos que surgen de la vibración en una nota musical concreta se producen en la misma nota a distancia de octava y en una nota a una distancia de 7 semitonos de forma ascendente.

Nota fundamental	1er Armónico	2º Armónico
<b>C</b>	C	G
<b>E</b>	E	B
<b>G</b>	G	D

*Tabla 3. Armónicos en Cmaj*

La componente de *D* (Re) se corresponde con el segundo armónico de la nota *G*, y la componente de *B* (Si) se corresponde con el segundo armónico que genera *E* (Mi). El segundo armónico que genera la nota *C* (Do), es *G* (Sol), que coincide con una nota del acorde.

Se puede profundizar esta información en [Recuero 1999] [Tipler 2007]

Este análisis corrobora la teoría de que, en términos frecuenciales, un acorde es necesario expresarlo como un vector de 12 componentes para tener en cuenta todas las influencias armónicas que produce cada instrumento.

#### 2.2.1.2 Modelos de acordes en trabajos previos.

Para realizar el reconocimiento de acordes en el cromograma, [Fujishima 1999] calcula el producto escalar entre vectores de 12 dimensiones que representan patrones de acordes predefinidos y el PCP derivado de la señal de audio. En cada inventariado, el patrón de acorde que devuelve el valor más alto es el elegido.

[Fujishima 1999] elige 27 clases, que se traducen, en el lenguaje musical, en los acordes de tres notas más comunes. Desafortunadamente, el resultado no fue del todo favorable en el reconocimiento de acordes de las muestras de audio del mundo real.

[Harte and Sandler 2005] también utilizaron el producto vectorial entre los vectores del cromograma y plantillas de acordes, sólo que ellos ampliaron el número de clases a 48 (Triadas de modo mayor, menor, aumentado y disminuido).

En el trabajo de [Bello and Pickens 2005], cada acorde se modeló como una distribución gaussiana de 12 dimensiones. En la cual, la media de la dimensión correspondiente a la nota del acorde se establece como 1, y las demás a 0. Además se asume una alta covarianza entre las notas de los acordes. Los vectores de chroma se normalizan con valores del intervalo [0, 1]. Otros trabajos similares han sido los de [Catteau et al. 2007] y [Papadopoulos and Peeters 2007].

El modelo de utilizar plantillas teóricas para los acordes es un desarrollo teórico-musical. La principal ventaja de esta forma es que los datos son independientes. Sin embargo, no se utilizan las propiedades físicas de muestras reales.

En el trabajo de [Sheh and Ellis 2003] también utilizan el modelo del cromograma para representar la señal, y asumen la distribución de gaussiana multidimensional para cada clase o tipo de acorde. Sin embargo, esas plantillas de acordes serán aprendidas con datos reales.

Como conclusión, se puede observar que en los trabajos conocidos sobre el reconocimiento de acordes, se han utilizado numerosos enfoques a la hora de expresar un acorde en términos físicos. A pesar de que en el terreno músico-teórico los acordes se expresan como combinación de notas individuales, en el terreno físico-teórico los acordes son expresados como un vector de 12 componentes. Esto facilita el reconocimiento de acordes sobre datos reales de una determinada transformación del espectro musical, llamada cromograma.

### 2.2.2 Alto nivel:

Una de las técnicas más adoptadas para la detección de acordes en audio es la detección analítica.

#### 2.2.2.1 Detección analítica.

La detección analítica tiene como objetivo desarrollar técnicas que permitan identificar correctamente, a partir de una serie de parámetros, un suceso desconocido. Para obtener los parámetros que influyen en esa decisión, es necesario un entrenamiento previo. Este entrenamiento se lleva a cabo usando sucesos previos ya conocidos. Los parámetros normalmente se traducen en patrones estadísticos.

Si se consideran los sucesos como fragmentos de audio, y se calculan los parámetros necesarios a fin de detectar acordes en dichos sucesos, se puede modelar como un problema de detección analítica.

Concretamente, los sucesos u **observaciones** serán los vectores de cromograma correspondientes a fragmentos de audio. Las alternativas que se disponen para decidir, también llamadas **clases**, serán cada uno de los acordes.

El entrenamiento se llevará a cabo con fragmentos de audio asociados a una clase ya conocida. El proceso de asignar la clase correcta a cada fragmento se conoce como **etiquetado**.

El etiquetado es necesario que se haga bajo la supervisión de un experto musical que sea capaz de averiguar el acorde que está sonando en cada momento.

La Figura 8 se describe el proceso de etiquetado que se sigue en este caso.

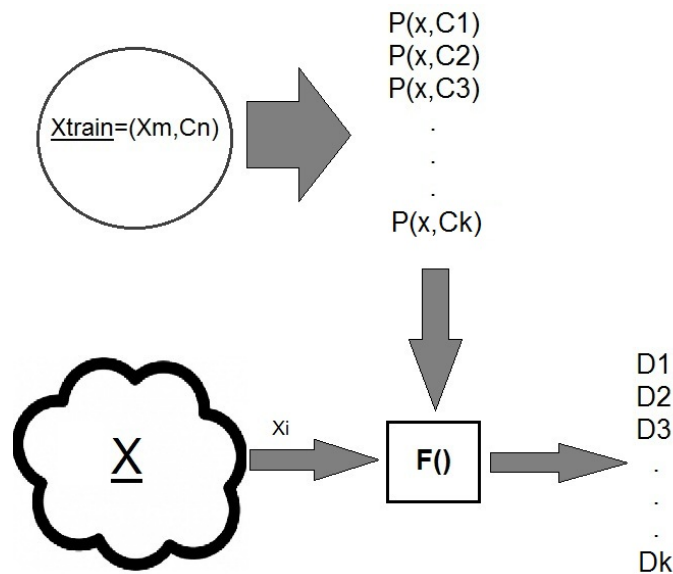


Figura 8. Proceso de etiquetado

- **C1,C2,...,Ck** representan cada una de las clases.
- **Xtrain=(Xm,Cn)** son las observaciones etiquetadas que se utilizan para entrenar el detector.
- Los parámetros que se obtienen al entrenar el detector se utilizan para calcular la probabilidad conjunta de cada una de las clases y el espacio muestral **x**.  
**P(x,C1),P(x,C2),...**

- $\underline{X}$  es el espacio de observación, es decir, las muestras que pasarán por el detector.
- $F()$  es la función de detección que se aplica a  $\underline{X}$ .
- $X_i$  es una determinada observación que pasará por la función de detección.
- $D_1, D_2, \dots, D_k$  representan cada una de las posibilidades de detección. En este caso se corresponden con las clases.

En este trabajo se va a emplear un **estimador de máxima verosimilitud**, también conocido como *estimador ML*.

El estimador ML es un método habitual para encontrar los parámetros de un modelo determinado.

Se utiliza bajo los supuestos siguientes:

- Las observaciones son **independientes**.
- Equiprobabilidad** entre las probabilidades a priori de las clases:  $(P(C_1), P(C_2), \dots)$ .

La detección del acorde  $C_k$  para la muestra  $X_i$ :

$$k = \arg \left\{ \max_g \left\{ \int_{X_i} P(X, C_k) \right\} \right\}$$

Según el teorema de Bayes:

$$k = \arg \left\{ \max_g \left\{ \int_{X_i} P(X/C_k) * P(C_k) \right\} \right\}$$

De acuerdo a la propiedad de equiprobabilidad, se puede simplificar como:

$$k = \arg \left\{ \max_g \left\{ \int_{X_i} P(X/C_k) \right\} \right\}$$

Si  $P(x/ C_k)$  se supone una función gaussiana; para calcularla se debe obtener, a partir del conjunto de entrenamiento, la media y varianza de cada clase.



### 2.2.2.2 Detección de acordes en trabajos previos.

La variedad de las posibles combinaciones de notas que surgen durante una franja de tiempo es extensa. Esto hace el proceso de estimación del acorde entre todo ese conjunto de notas una tarea difícil.

Como se ha indicado en el apartado 2.2.1 los acordes en la música están acompañados por otros elementos que dificultan su reconocimiento.

Bajo la premisa de que la capacidad cognitiva humana provoca que un acorde en música mantenga una duración por encima de un mínimo de tiempo, existen estrategias para eliminar desviaciones de corta duración con el fin de evitar el reconocimiento de falsos cambios de acorde.

A continuación se dan ejemplos de detectores más sofisticados de otros autores que se desarrollan bajo estos supuestos.

[Temperley and Sleator 1999] emplean el algoritmo de Viterbi, que permite encontrar la secuencia de datos más probable en un modelo oculto de Markov, para obtener transiciones suaves entre acordes.

[Fujishima 1999] detecta los límites de acorde introduciendo un procedimiento llamado “chord change sensing”(detección de cambio de acorde), que determina el punto de cambio de un acorde por la monitorización de la dirección de los vectores del cromograma.

La forma más adoptada hasta el momento para obtener secuencias de acordes más suavizadas es la de usar modelos probabilísticos de series temporales (“probabilistic time-series model”). [Raphael and Stoddard 2003], [Sheh and Ellis 2003],...

El modelo predominante es el de los modelos ocultos de Markov (HMM), conocido por ser una herramienta muy efectiva en el reconocimiento del habla, ya que los HMMs están orientados al reconocimiento de formas temporales y no solapan eventos en el tiempo. La sucesión de acordes en la música sigue ese mismo paradigma. Existe más información en: [Mauch 2010]

### 2.2.3 Métodos de evaluación de la detección.

El método más inmediato para evaluar la calidad es el cálculo del **porcentaje de error** de entre un número de muestras detectadas y observar la influencia que tienen en éste las variaciones de determinados parámetros que intervienen en la detección.

Otro de los métodos más utilizados es la **matriz de confusión**.

Se define como una matriz cuadrada de  $n$  dimensiones donde  $n$  representa el número de clases. Dicha matriz muestra la relación entre dos series de medidas correspondientes al área en estudio. La primera serie corresponde a datos reales observados y representan las filas. La segunda corresponde a la clasificación obtenida del reconocedor y representan las columnas.

Esta matriz es muy útil a la hora de reconocer si hay confusión entre dos tipos de acordes y en qué medida.

	<i>Cmaj</i>	<i>Cmin</i>	<i>Dbmaj</i>
<i>Cmaj</i>	15	10	0
<i>Cmin</i>	5	18	2
<i>Dbmaj</i>	0	0	25

*Tabla 4. Matriz de confusión*

Por ejemplo, la matriz de confusión de la Tabla 3, indica lo siguiente:

- De los 25 fragmentos de *Cmaj*, se han reconocido correctamente 15 y se han estimado erróneamente 10 en forma de *Cmin*.
- De los 25 fragmentos de *Cmin*, se han reconocido correctamente 18 y se han estimado erróneamente 5 en forma de *Cmaj* y 2 en forma de *Dbmaj*.
- Todos los fragmentos de *Dbmaj* se han estimado correctamente.

## 2.3 Aplicaciones comerciales

Hay numerosas aplicaciones en el mercado que tienen que ver con la detección de melodía y acordes a partir del audio. Aquí se exponen algunas de las más importantes.

Una herramienta para la transcripción de melodía o notas monofónicas es el programa *Melodyne* (<http://celemony.com>) (Figura 7). Pero tal y como detallan en su página web, “el problema con el material polifónico es que *Melodyne* plugin no tiene manera de aislar de forma individual las notas que forman cada acorde”, es incapaz de reconocer acordes.

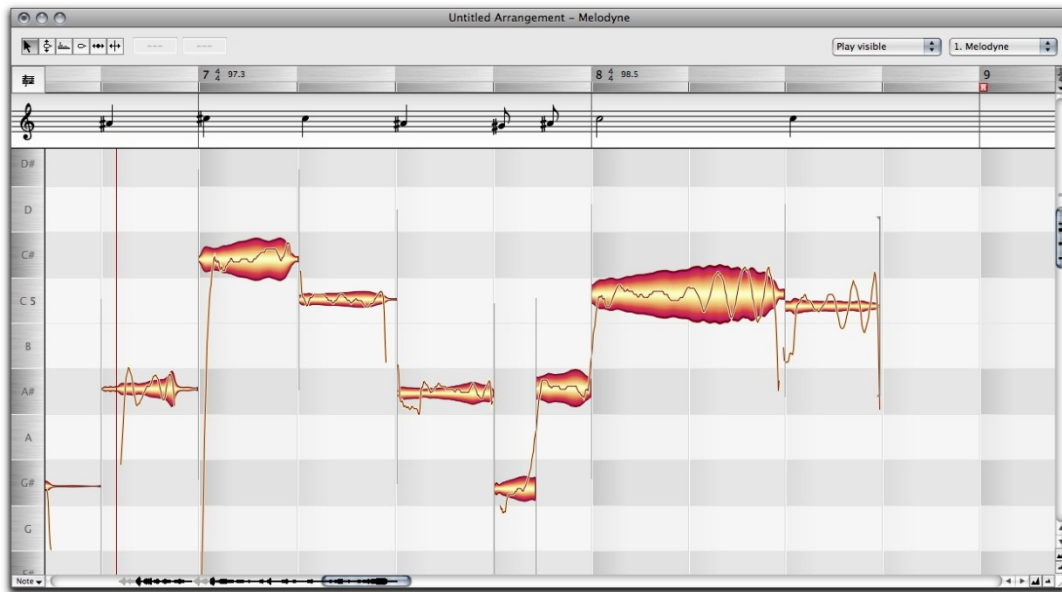


Figura 9. Melodyne.

Otra aplicación que guarda relación, pero no es exactamente un reconocedor de acordes, es el afinador polifónico de la empresa TC Electronic, llamado *TC electronic polytune* (Figura 8). Un afinador de guitarra, originalmente funciona de forma monofónica, es decir, capta el sonido que produce una cuerda determinada al tocarse y devuelve información sobre la frecuencia a la que se encuentra y la distancia a la que está de la frecuencia de afinación del instrumento, con el fin de dar información al músico sobre si debe tensar o destensar la cuerda para conseguir la frecuencia deseada. Este afinador permite que el guitarrista toque todas las cuerdas simultáneamente y detecte qué cuerda de la guitarra está desafinada y en qué grado, con el fin de ahorrar tiempo.



Figura 10. Polytune

*Chord Pickout* (Figura 11) y el plugin *Chordino* (Figura 12) del programa *NNLS Chroma* reconocen acordes de la señal de audio de forma bastante fiable. La interfaz en los dos casos presenta cómo los acordes se van sucediendo a medida que la muestra de audio en forma de onda va avanzando. Es una forma muy intuitiva pero no es directamente extrapolable a la transcripción musical.

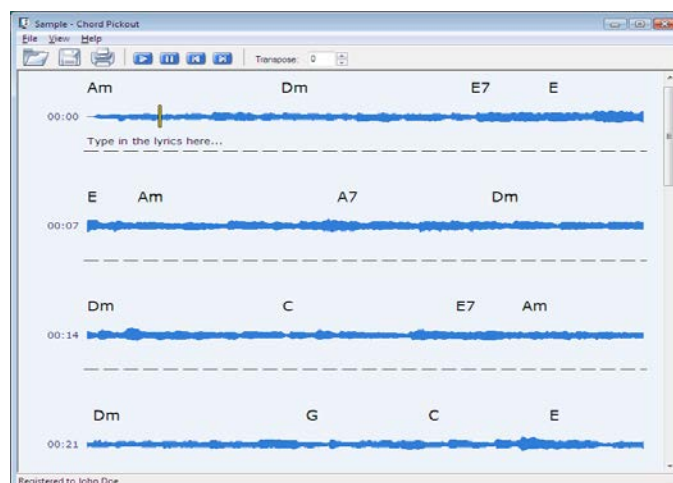


Figura 11. *Chord Pickout*

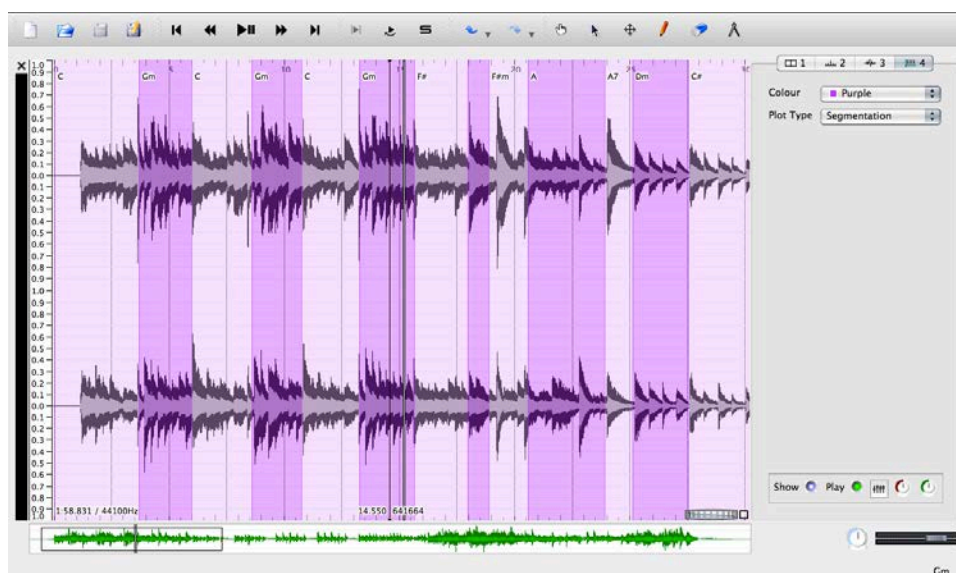


Figura 12. *Chordino*

El plugin *Chord Wizard* del programa *Band in a box* (Figura 13) aporta a los programas descritos antes, la transcripción de los acordes a una notación musical simple. Divide la canción en compases y distribuye los acordes reconocidos en éstos, permitiendo al músico su lectura inmediata.

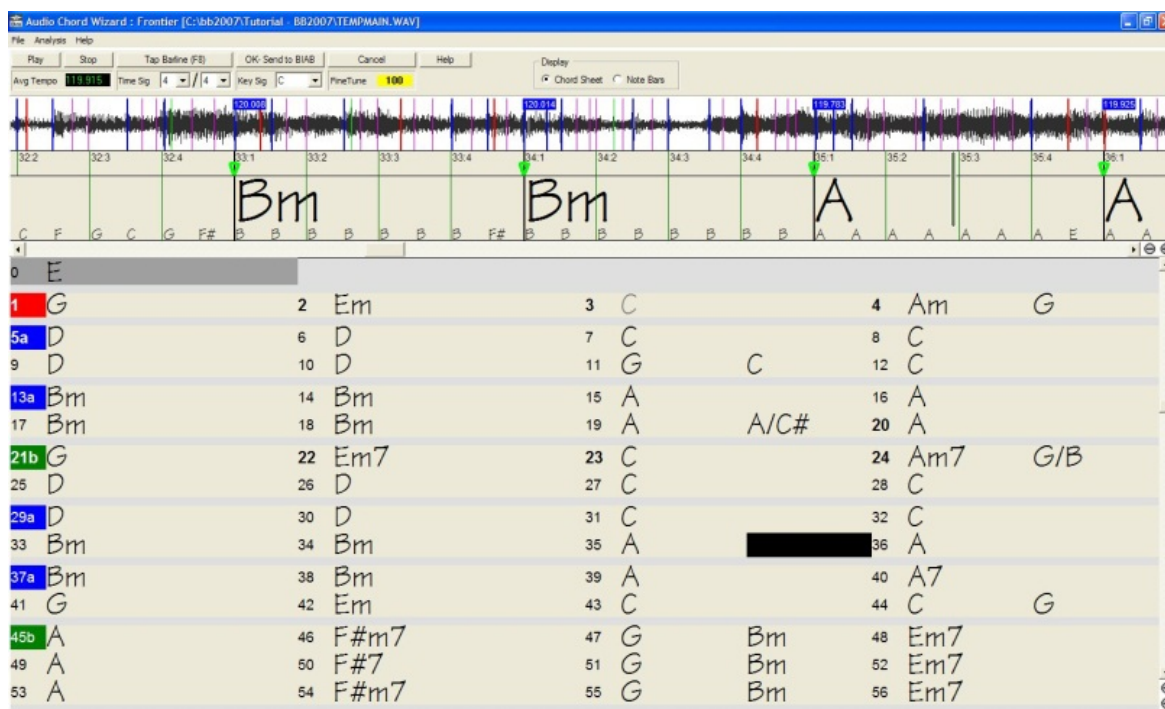


Figura 13. Band in a box

A pesar de los avances descritos, todavía no se ha conseguido una transcripción perfecta de los acordes de determinadas canciones, por lo que sigue siendo un campo en desarrollo.



## **Capítulo 3. Diseño e implementación**

### 3.1 Introducción

La propuesta, como se ha descrito antes, consiste en crear un prototipo que sea capaz de reconocer los acordes que se van sucediendo en una muestra de audio determinada.

La forma que se ha adoptado ha sido la de modelarlo como un problema de decisión estadística.

Como tal, primero es necesario crear una base de datos consistente de muestras de audio debidamente etiquetadas.

Las **observaciones** serán fragmentos de las muestras que componen la base de datos.

Las **clases** que se utilizarán en el etiquetado y en el reconocimiento hacen un total de 25. 24 tipos de acordes, que se corresponden con las triadas mayores y menores de cada una de las 12 notas y una clase para los casos donde no suene acorde ninguno. En este primer punto se utilizará una aplicación para extraer la información de cromograma de los archivos de audio: *Sonic Visualiser*.

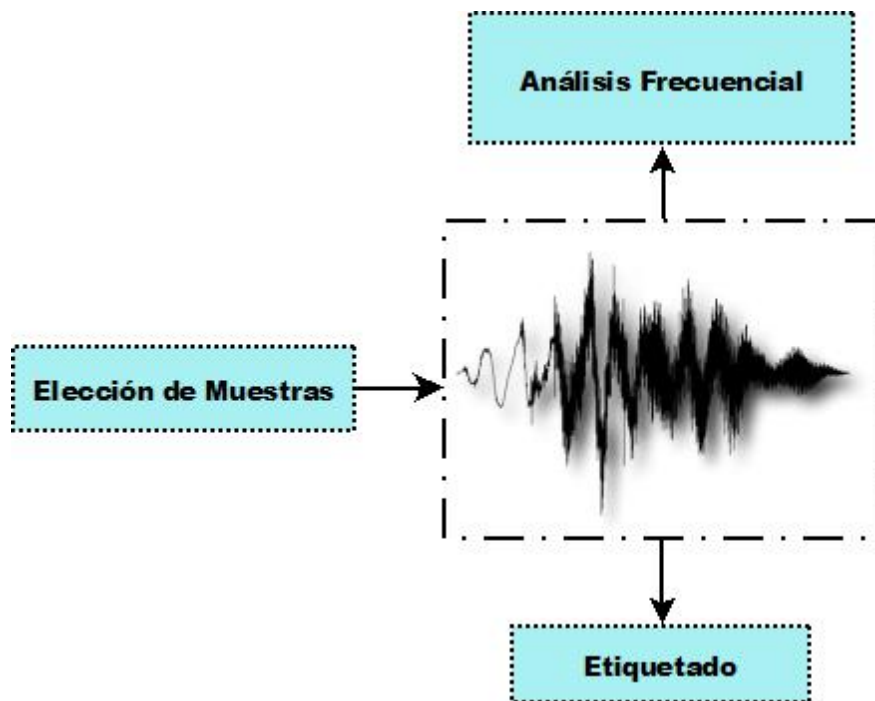
Una vez dispuesta la base de datos, se diseñará un prototipo que, a partir de esas muestras, actúe como detector de los acordes que las componen. Se utilizará el software matemático *Matlab* para su desarrollo.

### 3.2 Creación de la base de datos.

Este apartado tiene como finalidad obtener el conjunto de observaciones para entrenar y validar el reconocedor.

Se ha dividido en tres partes:

- Primero, se **eligen las muestras de audio** que van a formar la base de datos.
- Cada muestra pasará por un **anализador de frecuencias** que fragmentará el audio y obtendrá su cromograma (Descrito en el apartado 2.2.1.1)
- A su vez, cada muestra será **etiquetada**, por un experto en conocimientos musicales, con los distintos acordes que la conforman.



*Figura 14. Esquema de creación de la base de datos*

#### 3.2.1 Elección de las muestras de audio:

En la base de datos es necesario que exista información clasificada en dos grandes grupos, que tienen que ver con la forma en la que pueden aparecer los acordes en la música. Un primer grupo de acordes aislados, es decir, sin influencias de otros elementos, y un segundo grupo de acordes en canciones.



### 3.2.1.1 Acordes aislados:

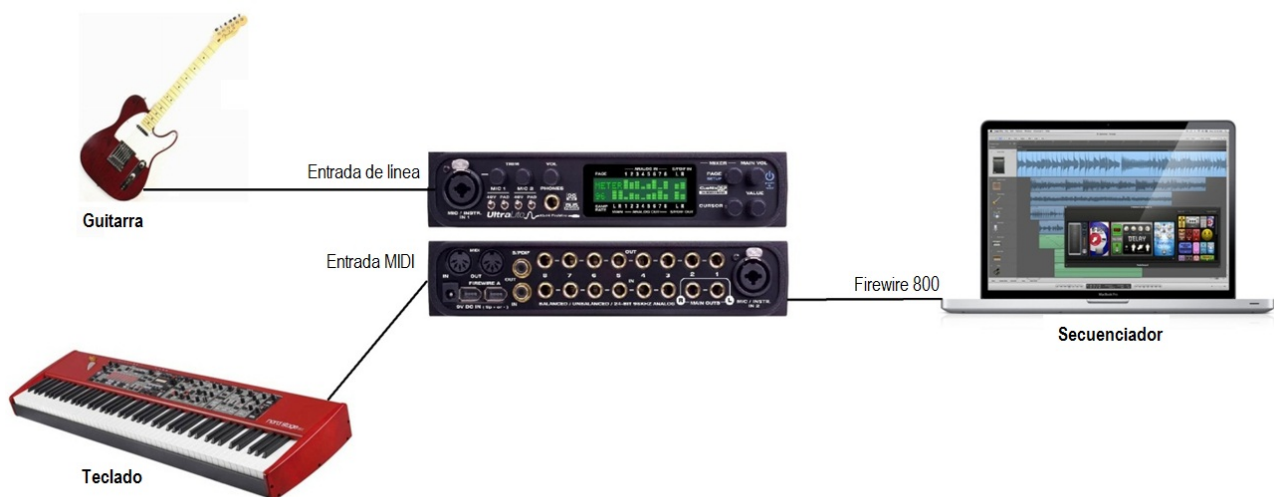
Se necesitan, en primer lugar, muestras de acordes para inicializar el entrenamiento del reconocedor. Los acordes de estas muestras van a estar aislados, es decir, no van a tener influencias procedentes de otras fuentes de audio que no estén ejecutando ese acorde en particular. (Batería, voz, etc.)

Además, las muestras se grabarán con distintos tipos de instrumentos. Cada uno generará una serie de armónicos, característicos del timbre y de la forma de ejecución del acorde que aportarán variedad a la base de datos.

Se han elegido tres instrumentos. Un **piano acústico**, un **piano Rhodes** y una **guitarra**. Son instrumentos muy presentes en la música popular del último siglo y frecuentemente utilizados en las grabaciones para hacer sonar acordes.

El piano Rhodes es un tipo de piano electromecánico cuyo sonido surge de la acción de un martillo al golpear un diapason. Posee un timbre distinto a cualquier piano convencional, que sustituye el diapason por un entramado de cuerdas.

A continuación se presenta un esquema de interconexión del equipo utilizado para la grabación de las muestras (Figura 13).



*Figura 15. Esquema de interconexión*

El proceso de grabación de los dos pianos ha sido mediante el envío de mensajes MIDI<sup>2</sup> desde el teclado a una tarjeta de sonido. El programa encargado de interpretar esos mensajes en forma de notas musicales se denomina secuenciador.

A su vez, este secuenciador gestiona librerías de grabaciones reales de distintos instrumentos, de tal forma que, al reproducir los mensajes MIDI, se genera una muestra de audio como si se hubiese tocado con el instrumento elegido.

Si se desea profundizar en el tema del lenguaje MIDI, puede consultarse: [Sala 1997]

Mientras que en el piano Rhodes se han grabado las notas del acorde tocadas simultáneamente, en el piano acústico se han grabado las notas del acorde de forma arpegiada<sup>3</sup> en diferentes octavas con el fin de imitar otro tipo de ejecución de acordes dado en la música.

La **guitarra** ha sido grabada directamente conectándola a la entrada de línea de la tarjeta de sonido mediante un cable con conector *jack*. La tarjeta amplifica esa señal de entrada, la digitaliza y la envía al secuenciador, que es el encargado de grabar en tiempo real.

En la guitarra, como en el piano Rhodes, las notas del acorde han sido tocadas simultáneamente.

Con cada instrumento se ha grabado una muestra de 30 segundos de duración de cada uno de los 24 tipos de acorde. Desde el secuenciador se exporta cada una de las muestras en un formato de audio digital (mp3).

### 3.2.1.2 Canciones:

Además de las muestras grabadas, en la base de datos se van a incluir canciones.

En éstas, los instrumentos que ejecutan acordes están acompañados de otro tipo de elementos sonoros que, en ocasiones, ejercen una función enmascaradora de los primeros.

---

<sup>2</sup> MIDI: protocolo de comunicación estándar para la comunicación a tiempo real entre computador y teclado musical, entre otros.

<sup>3</sup> Arpeggio: Forma de ejecutar los tonos de un acorde. En lugar de tocarlos de manera simultánea, se hacen oír en sucesión rápida.

Se han escogido temas de dos intérpretes muy representativos de la música popular.

a) **Beatles.**

Por ser uno de los grupos que más influencia han ejercido en la música actual.

Se ha podido aprovechar una base de datos ya etiquetada que se encuentra en la página: [isophonics Beatles]

- Canciones: "Come together", "Help", "I should have known better", "I need you", "And I love her", "Let it be", "Tell me why", "Yellow submarine", "Here comes the sun", "Any time at all", "I'll cry instead", "Things we said today", "In my life", "Mean mr. Mustard", "Polythene pam", "I'll be back" y "Twist and shout".

b) **Michael Jackson.**

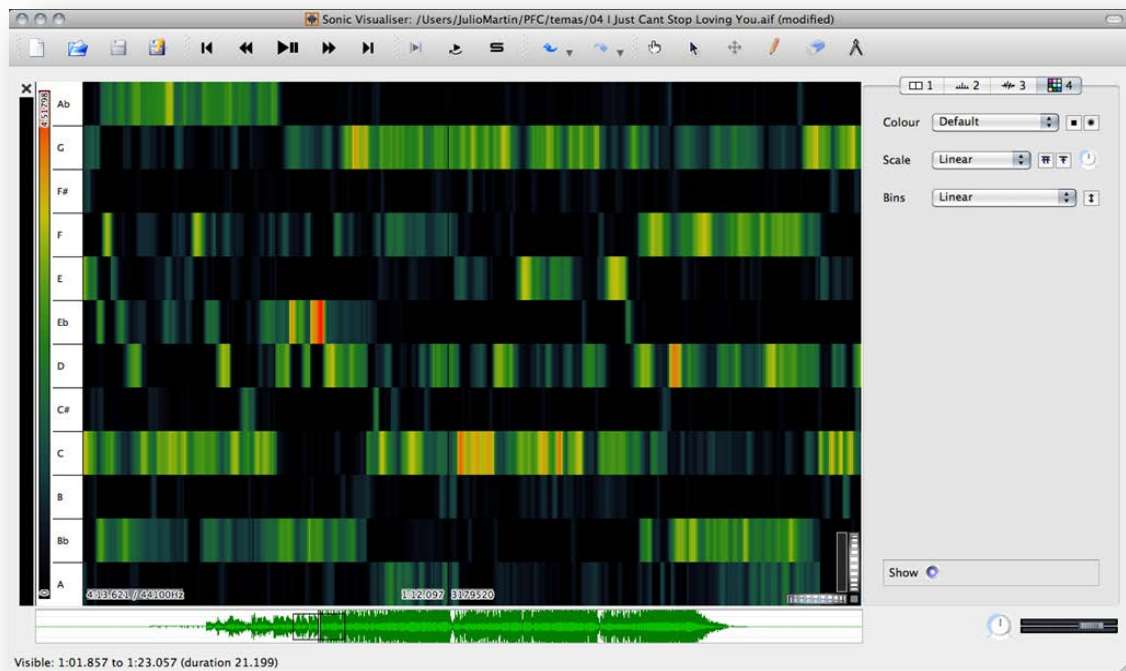
Además de influir al igual que los Beatles en la música popular, ha tenido una de las trayectorias profesionales más longevas. Esto permite elegir temas de diferentes géneros musicales (*Rock, Funky, Pop*) con una notable diferencia al estilo de los Beatles.

- Canciones: "Earth song", "Heal the world", "Wanna be starting something", "You are not alone", "Thriller", "Beat it", "I just can't stop loving you", "Blame it on the boogie", "Black or white", "Ben" y "Rockin' robin".

### 3.2.2 Análisis frecuencial de las muestras.

Para ello se ha utilizado un complemento (plugin) llamado *NNLS Chroma* del programa *Sonic Visualizer*, cuya principal función es extraer el cromograma de una muestra de audio determinada.

#### 3.2.2.1 Analizador. NNLS Chroma



*Figura 16. Cromograma de una muestra de audio*

El programa nos permite exportar la información de cada cromograma en forma de matriz numérica de tamaño:  $N \times 12$ . Nos referiremos a ella como **matriz Song**.

En la matriz Song, las 12 columnas representan las doce notas musicales y el número indeterminado de filas son los fragmentos en los que ha sido muestreada la canción. Cada elemento de la matriz representa la potencia una determinada nota en un fragmento en concreto. Esta potencia está normalizada.

Fragmento/ semitonos	A	Bb	B	C	Db	D	Eb	E	F	Gb	G	Ab
1	0.234	0.044	0.565	0.756	1.123	0.987	0.002	0.336	1.456	0.233	0.435	1.78
2	1.952	0	0.012	0.119	2.156	0	0.003	0.048	1.306	0.238	0	0.2637
...	...	...	...	...	...	...	...	...	...	...	...	...
n	(n,A)	(n,Bb)	(n,B)	(n,C)	(n,Db)	(n,D)	(n,Eb)	(n,E)	(n,F)	(n,Gb)	(n,G)	(n,Ab)

Tabla 5. Ejemplo de Matriz Song

### 3.2.3 Etiquetado de muestras.

Todas las muestras de la base de datos tienen que ser etiquetadas con los acordes correspondientes.

Cabe mencionar, que las clases con las que se trabaja en esta propuesta (25), representan una parte de los acordes de tres notas (triadas). Éstos son un subconjunto de todos los que existen en la música (cuatriadas, quintiadas, etc.). La solución es adaptar el etiquetado de acordes a esta limitación. Hay determinados acordes de 4 y 5 notas que han tenido que sustituirse por un acorde reducido de éstos que contiene 3 notas (método descrito en la siguiente página).

Esto será un elemento a tener en cuenta en la evaluación de la detección.

Para expresar las etiquetas se ha creado una matriz de etiquetado por cada muestra de audio. Esa matriz va a ser conocida como **matriz Chord**.

La matriz Chord representará los cambios de acorde que se van sucediendo en la muestra de audio correspondiente.

Consta de tres columnas. La primera va a representar el momento temporal en el que se inicia un acorde ( $T_0$ ), la segunda es el momento en el que termina ( $T_f$ ), y la tercera el acorde correspondiente a ese momento temporal.

Las filas serán tantas como cambios de acorde existan en la canción.

Fragmento/acorde	$T_0$	$T_f$	Acorde
1	0.000	4.569	Amin
2	4.569	9.254	Bbmaj
...	...	...	...
m	(m, $T_0$ )	(m, $T_f$ )	(m,Acorde)

Tabla 6. Ejemplo de Matriz Chord

A continuación se describen los pasos que se han seguido para obtener los acordes y su duración en cada una de las canciones.

- a) **Identificación de los acordes en la muestra:** Previamente a la creación de la matriz de etiquetado es necesario que un experto musical escuche e identifique los acordes que están sonando.

Para agilizar el proceso, puede ayudarse de una partitura de la canción.

- b) **Adaptación de los acordes:** Una vez identificados, es necesario una adaptación de todos los acordes que suenan al subgrupo de acordes que se han tenido en cuenta para el proyecto.

Cualquier conjunto de notas de todo acorde mayor de tres notas contiene una triada con la que comparte tónica y que actúa como acorde reducido de éste.

Por ejemplo, la cuatriada *Cmaj7* está compuesta por: *Do* (tónica), *Mi*, *Sol* y *Si*. Este acorde se puede reducir a la triada con la que comparte tónica. Es decir, *Cmaj*: *Do* (tónica), *Mi* y *Sol*. Puede consultarse más información sobre tipos de acordes en: [Herrera 1984]

- c) **Construir la matriz de etiquetado:** El procedimiento común es el de reproducir la secuencia de audio y parar la reproducción en los momentos que se producen los cambios de acorde. Entonces se anota a mano cada momento y el acorde asociado. La principal desventaja es que la tarea es lenta y tediosa. La forma que se ha adoptado en este proyecto ha sido la de utilizar eventos MIDI para registrar los momentos temporales de cambio de acorde. El procedimiento es grabar a tiempo real, durante la reproducción de la canción, dichos mensajes desde un teclado musical. (La fiabilidad del MIDI en tiempo real es mayor que la de un ratón o teclado de ordenador utilizado para parar la reproducción en el procedimiento común).

El músico encargado debe tocar las tónicas de los acordes a medida que éstos vayan apareciendo. Al finalizar la canción, la información temporal y las tónicas de todos los mensajes emitidos se traslada a la matriz de etiquetado. Para finalizar, tan solo hay que sustituir las tónicas por el acorde.

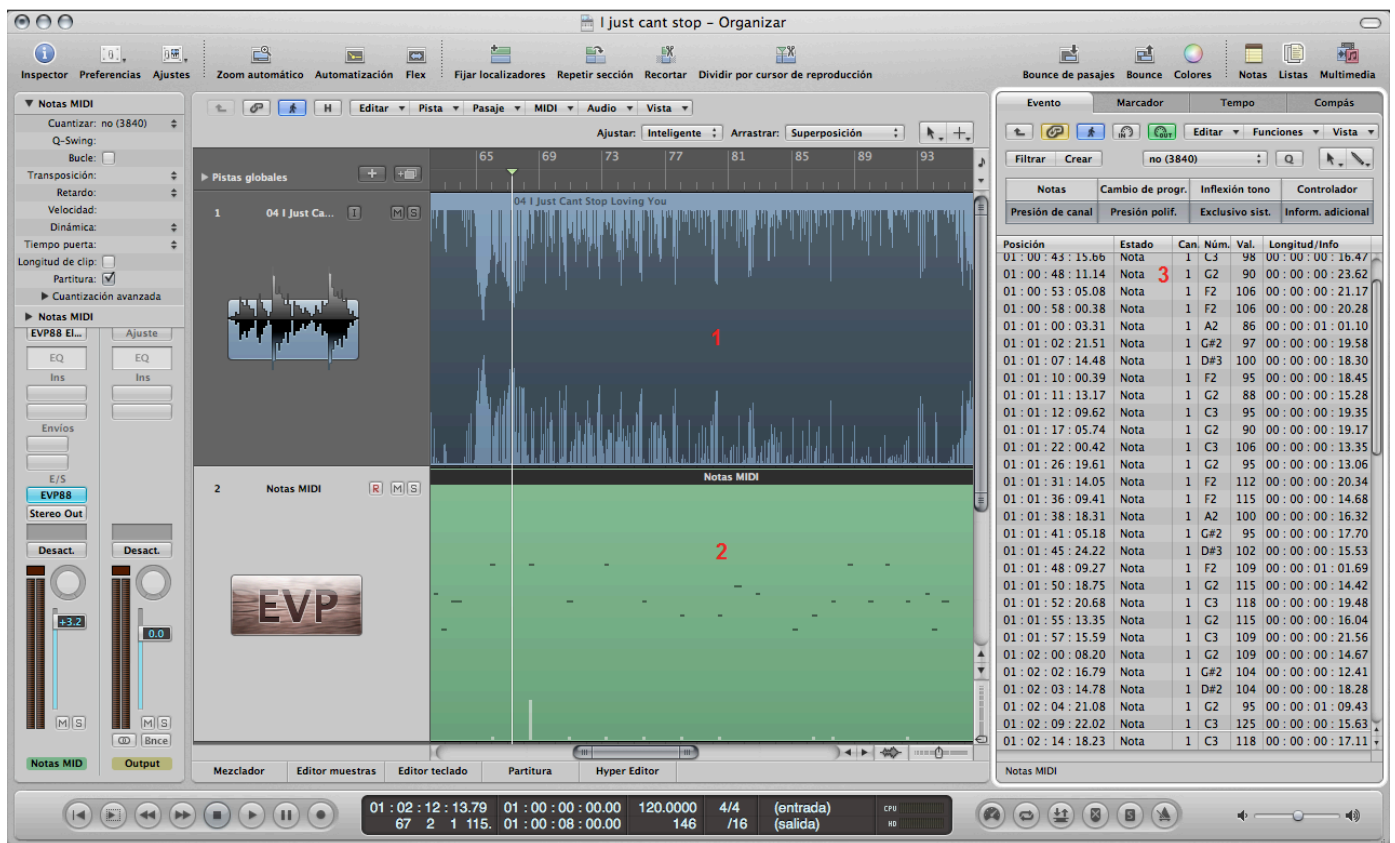
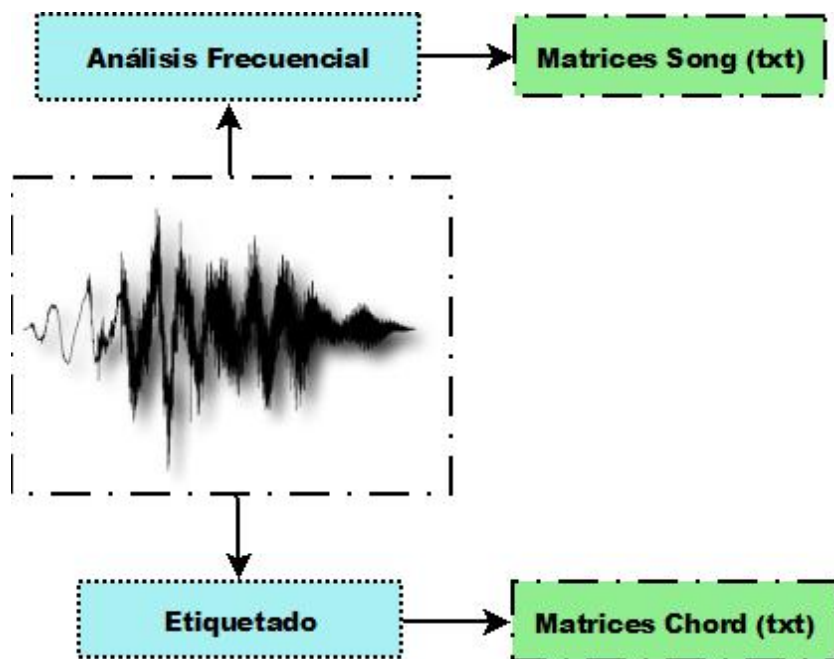


Figura 17. Etiquetado en el secuenciador.

En la Figura 17 se muestra una imagen del secuenciador. La onda sonora (1) está sincronizada con los datos MIDI (2). En la parte derecha (3), el secuenciador nos muestra la información de tiempo y nota asociada de cada evento MIDI.

Los temas que se han etiquetado bajo este procedimiento son: “Earth song”, “Heal the world”, “Wanna be starting something”, “You are not alone”, “Thriller”, “Beat it”, “I just can’t stop loving you”, “Blame it on the boogie”, “Black or white”, “Ben” y “Rockin’ Robin”. Todos ellos de Michael Jackson.

La Figura 18 muestra un esquema de la etapa de análisis de las muestras y el etiquetado.



*Figura 18. Esquema del análisis y etiquetado.*



### 3.3 Prototipo.

El prototipo consta de 4 partes diferenciadas:

- En primer lugar, se introduce la información de la base de datos al software matemático, *Matlab*, para crear el conjunto de observaciones.
- A partir de esas observaciones, se extrae el conjunto de entrenamiento
- En tercer lugar, se implementará el detector a partir de esos datos de entrenamiento y se hallarán los acordes de distintas muestras de audio.
- Por último, se crearán herramientas que permitan evaluar la calidad de la detección obtenida.

#### 3.3.1 Introducción de la base de datos.

Los archivos que contienen la información de la base de datos utilizada son las matrices *Song* y las matrices *Chord*. Este apartado describe la introducción de estas matrices, en formato *.txt*, al software matemático.

##### 3.3.1.1 Carga de matrices Song.

Las matrices *Song* se cargan directamente al software *Matlab* con el comando ***load.m***, ya que son numéricas en su totalidad.

##### 3.3.1.2 Carga de matrices Chord.

Sin embargo, en las matrices *Chords*, cada acorde está representado con una cadena de caracteres (*Strings*).

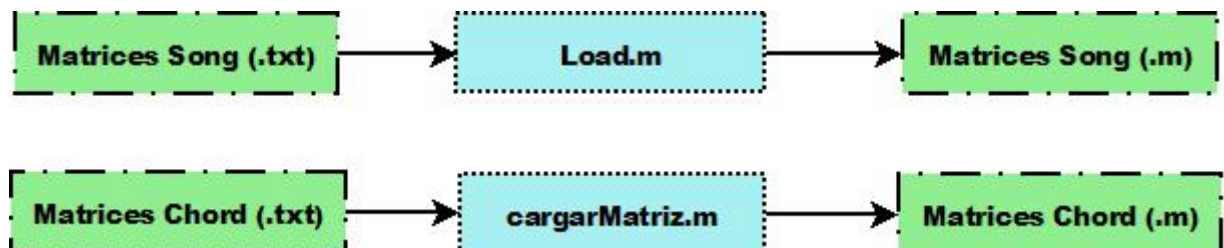
Para poder manejar los datos en *Matlab* tenemos que convertir estos acordes a números usando una asociación numérica. Como se ha indicado en el apartado 2.1.1, cada nota alterada puede nombrarse de dos formas. En este proyecto se utilizarán tan solo los acordes en formato bemol (*b*), para simplificar el proceso.

En la Tabla 7 se muestra la asociación numérica a cada uno de los acordes.

Acorde	Número
Amaj	1
Amin	2
Bbmaj	3
Bbmin	4
Bmaj	5
Bmin	6
Cmaj	7
Cmin	8
Dbmaj	9
Dbmin	10
Dmaj	11
Dmin	12
Ebmaj	13
Ebmin	14
Emaj	15
Emin	16
Fmaj	17
Fmin	18
Gbmaj	19
Gbmin	20
Gmaj	21
Gmin	22
Abmaj	23
Abmin	24
No acorde	25

*Tabla 7. Asociación numérica de los acordes*

Esta conversión se lleva a cabo con la función **cargarMatriz.m**. Figura 19.



*Figura 19. Esquema de la introducción de la base de datos*

### 3.3.1.3 Alineamiento de las matrices Song y Chord.

Este apartado consiste en incorporar las etiquetas de las matrices *Chord* a las filas de las matrices *Song*. Para ello, se debe calcular previamente la duración de cada fragmento de audio y compararlo con la duración de los acordes.

Para su cálculo se tendrán en cuenta determinados parámetros manejados por el programa *Sonic Visualizer*.

Los datos del inventariado al calcular la DFT de cada muestra de audio son:

*Tamaño de ventana: 16384 muestras*  
*Incremento de ventana: 2048 muestras.*

Los datos obtenidos al cargar la canción *You are not alone* son:

*Tema: "You are not alone"*  
*Duración: 6.01 min (361 sec)*  
*Muestras totales: 15982847*  
*Filas del espectrograma: 7739*

Aunque el tamaño de ventana es de 16384 muestras (unidad utilizada por el programa), al producirse solapamiento, la duración real del inventariado es de 2048 muestras.

La duración de la canción está expresada en muestras y en segundos.

Bajo la suposición de que el tamaño en muestras del inventariado coincide con el de cada fragmento, se puede obtener su duración en segundos.

$$(362 * 2048) / 15982847 = 0.04639 \text{ sec.}$$

Para comprobar la última suposición, se ha de multiplicar la duración obtenida por el número de fragmentos en los que ha sido dividida la muestra.

$$0.04639 * 7739 = 359 \text{ sec} \approx 361 \text{ sec.}$$

El último paso es comprobar que la duración de los fragmentos se cumple para todos los casos. Para ello se hacen los mismos cálculos en otra canción.

*Tema: "Black or white"*

*Duración: 3.19 min = 199 sec*

*Muestras totales: 8780543*

*Filas del espectrograma: 4261*

$$(199 * 2048) / 8780543 = 0.04641 \text{ sec}$$

$$0.04641 * 4261 = 198 \text{ sec} \approx 199 \text{ sec}$$

Queda verificado que la duración de cada fragmento coincide en los dos casos. El programa muestrea el audio aproximadamente cada **0.0464 sec**.

Una vez calculado este dato, se implementa la función **etiquetarSong.m**, que es la encargada de alinear las etiquetas con las observaciones. Figura 20.



*Figura 20. Alineamiento de matrices Chord y Song.*

Esta función añade una columna a la matriz *Song* con la información del acorde correspondiente a cada fragmento. En la Tabla 8 se muestra un ejemplo

Fragmento	A	Bb	B	C	Db	D	Eb	E	F	Gb	G	Ab	Acorde
1	0.234	0.044	0.565	0.756	1.123	0.987	0.002	0.336	1.456	0.233	0.435	1.78	9
2	0.25	0	0.012	0.119	2.156	0	0.003	0.048	1.306	0.238	0	1.97	9
n	...	...	...	...	...	...	...	...	...	...	...	...	...

*Tabla 8. Matriz Song etiquetada*

### 3.3.2 Extracción de datos de entrenamiento.

Ya se dispone de las muestras con las etiquetas colocadas correctamente.

El siguiente paso consiste en extraer, de todas las observaciones, un conjunto de datos de entrenamiento y agruparlos para entrenar el clasificador.

La función encargada de definir las observaciones que se van a utilizar como entrenamiento es **dameFilas.m**. Figura 21.

Al ejecutarla se le pasa como parámetros el número de filas de la matriz y el porcentaje de éstas que se desea utilizar como conjunto de entrenamiento.

La función nos devuelve un vector fila con índices aleatorios correspondientes con las filas que se utilizarán para crear la matriz de entrenamiento.

La matriz de entrenamiento, por lo tanto, contendrá un tanto por ciento de filas de la matriz *Song* aleatoriamente escogidas.



Figura 21. Extracción de datos de entrenamiento.

#### 3.3.2.1 Agrupación de datos de entrenamiento.

Una vez que se tiene el conjunto de entrenamiento deseado, es necesario agrupar y clasificar los fragmentos de todas las matrices de entrenamiento en un vector de 25 matrices llamado *MatrizTrain*{}.

*MatrizTrain*={*Amaj()*,*Amin()*,*Bbmaj()*,*Bbmin()*,*Bmaj()*,*Bmin()*,*Cmaj()*,*Cmin()*,*Dbmaj()*,*Dbmin()*,  
*Dmaj()*,*Dmin()*,*Ebmaj()*,*Ebmin()*,*Emaj()*,*Emin()*,*Fmaj()*,*Fmin()*,*Gbmaj()*,*Gbmin()*,*Gmaj()*,*Gmin()*,  
*Abmaj()*,*Abmin()*}

La función que se encarga de llevarlo a cabo es **separaTrain.m**. Figura 22.

#### 3.3.2.2 Cálculo de medias y varianzas.

Para el entrenamiento del detector, se necesitan las medias y varianzas de cada una de las clases.

Se han creado dos Script<sup>4</sup> para hallar estos cálculos.

**calcularMedias.m** y **calcularVarianzas.m** calculan las medias y varianzas de cada una de las matrices del vector *MatrizTrain()*, y las agrupan en un par de vectores. Figura 22.

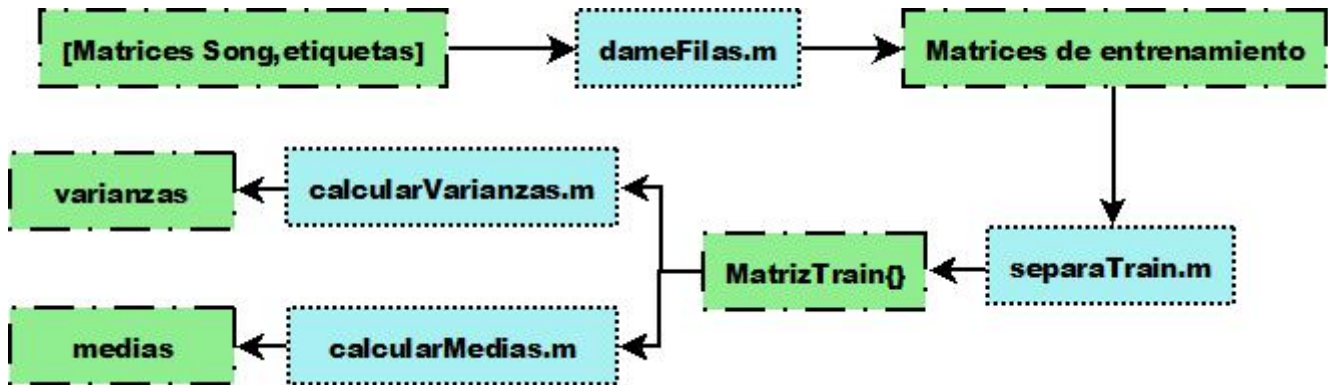


Figura 22. Agrupación de datos de entrenamiento

### 3.3.3 Detector.

En este bloque, se procederá a obtener los acordes que componen cada una de las muestras de audio de la base de datos. Primero, mediante un detector, se hallarán los acordes de cada uno de los fragmentos y, posteriormente, un algoritmo de postprocesado suavizará los cambios bruscos entre acordes teniendo en cuenta una serie de parámetros.

#### 3.3.3.1 Detector ML

Para implementar el detector se han supuesto las verosimilitudes como gaussianas y las probabilidades a priori de cada una de las hipótesis como equiprobables.

Bajo esta premisa, el decisor que se utilizará será el de **Máxima Verosimilitud**.

La función encargada de hacer esta detección es: **detectorML(x,M,V)**. Figura 23.

Los parámetros de entrada de la función son:

- **x**: Matriz *Song* para detectar.

---

<sup>4</sup> Script: En informática, se refiere a un archivo de texto en el que se almacenan órdenes secuenciales a un programa.

- **M**: Vector de medias.
- **V**: Vector de varianzas.

Para cada observación se calculan las 25 verosimilitudes:  $P(x/A_{maj})$ ,  $P(x/A_{min})$ , etc.

La función encargada en calcular cada una es **calcularGaussiana.m**. Figura 23.

Por último, se escoge la clase cuya verosimilitud sea mayor.

Cuando se tienen estimados la totalidad de los fragmentos de la muestra de audio, se devuelve la *matriz Song* con una última columna que representa los acordes estimados.

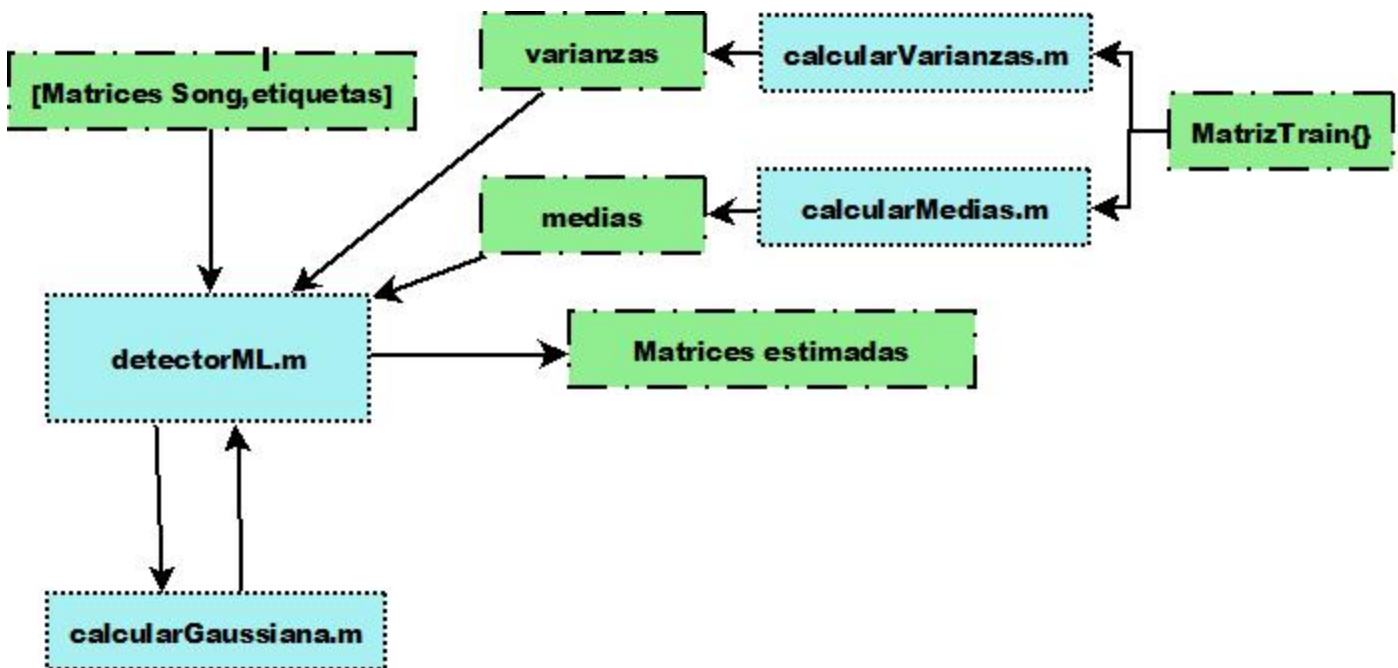


Figura 23. Esquema del Detector

### 3.3.3.2 Postprocesado.

El principal problema a la hora de reconocer acordes con éste método, es que no se tienen en cuenta las muestras colindantes de un fragmento determinado.

Se supone el caso de que, en 10 fragmentos sucesivos de la misma muestra de audio, se detectaran los acordes que se muestran en la Tabla 9:

Amaj	Amaj	Amaj	Dbmin	Amaj	Amaj	Amaj	Bmin	Bmin	Bmin
------	------	------	-------	------	------	------	------	------	------

*Tabla 9. Ejemplo de una secuencia de acordes.*

Bajo la premisa de que un acorde es imposible de ejecutar (y de ser reconocido) por un ser humano en un fragmento tan corto de tiempo (el muestreo se produce aproximadamente cada 0.05 sec), lo más probable es que el acorde *Dbmin* esté incorrectamente estimado y en su lugar, casi con toda probabilidad, coincidiría con el valor de las muestras colindantes a éste. Es decir, *Amaj*.

El error en la detección del acorde puede deberse a determinados factores presentes en la música que pueden excitar frecuencias que no pertenecerían al acorde en sí, y hacer que la detección sea errónea. Por ejemplo, un golpe a un plato de la batería excita un rango de frecuencias determinadas que pueden enmascarar una parte importante de la información del acorde.

La función encargada de llevar a cabo el postprocesado es ***suavizarEstimacion.m***. Figura 24.



*Figura 24. Esquema de postprocesado*

El cometido de esta función es eliminar los cambios repentinos que se producen por un error en la detección del acorde.

Para ello, una vez agrupados los fragmentos estimados, la nueva detección será el resultado de calcular la moda de un número determinado de muestras adyacentes del fragmento que queremos detectar.

El número de muestras se puede calcular con el *tamaño de ventana*. Parámetro que se introduce a la función cada vez que se ejecuta.



### 3.3.4 Evaluación de la detección

#### 3.3.4.1 Probabilidad de error

La función **probError.m** compara el vector de acordes estimados con el vector de etiquetas.

Calcula el porcentaje de acordes estimados erróneamente entre la totalidad de estimaciones.

#### 3.3.4.2 Matriz de Confusión.

Esta función calcula la matriz de confusión de cada muestra de audio estimada.

Cada columna de la matriz representa el número de predicciones de cada clase, mientras que cada fila representa a las instancias en la clase real.

La principal ventaja de la matriz de confusión es observar si se están confundiendo dos o más clases en la detección.

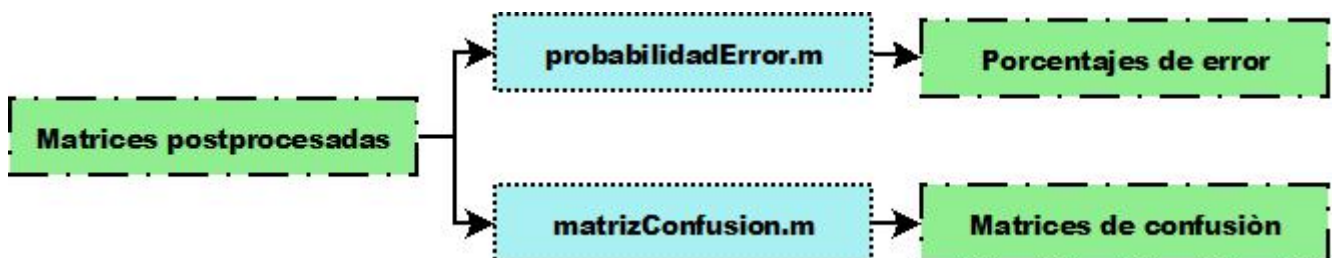


Figura 25. Evaluación de la detección.



## **Capítulo 4. Valoración experimental**

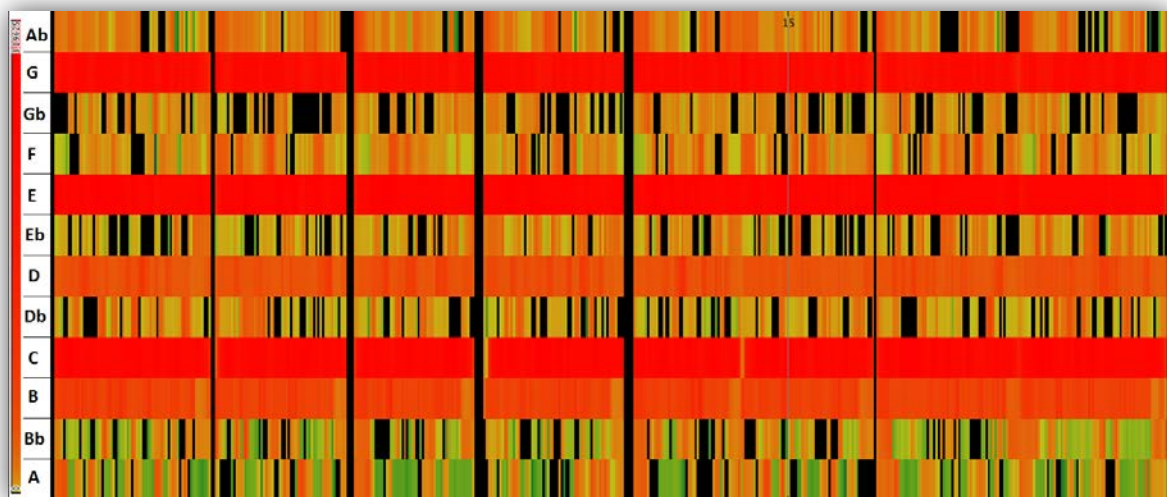
En este capítulo se evaluará la calidad del detector con ayuda de diversas herramientas. Además, se estudia la influencia de determinadas variables en la detección.

Los apartados en los que está dividido, son:

- **Análisis frecuencial** de las muestras de audio que componen la base de datos mediante la obtención de su cromograma.
- **Evaluación de la detección** de acordes aislados y de acordes en canciones mediante el porcentaje de error.
- **Influencia del conjunto de entrenamiento** y cálculo del porcentaje de entrenamiento óptimo.
- **Influencia del postprocesado** y cálculo del tamaño de ventana óptimo.
- **Cálculo de matrices de confusión** en la detección de acordes de las muestras de audio en función de distintos conjuntos de entrenamiento.

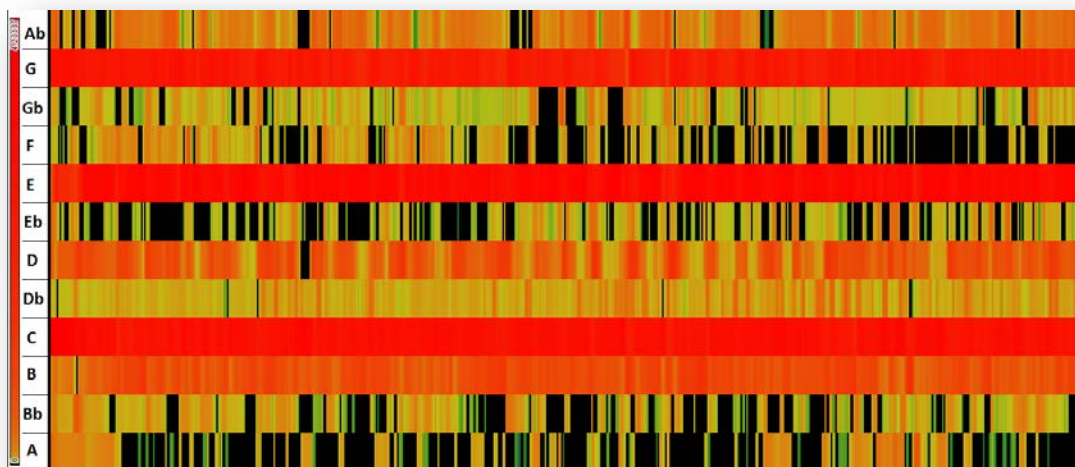
## 4.1 Cromograma de muestras.

Antes de presentar el resultado de la detección, con la finalidad de conocer las características de las muestras de audio de las que se dispone, se mostrará el cromograma de cuatro muestras de audio diferentes. Los primeros cromogramas (Figuras 26, 27 y 28) representan el acorde de Cmaj tocado por tres instrumentos diferentes. El último (Figura 29) muestra el tema “Billie Jean” de Michael Jackson.



*Figura 26. Cmaj tocado en un piano Rhodes.*

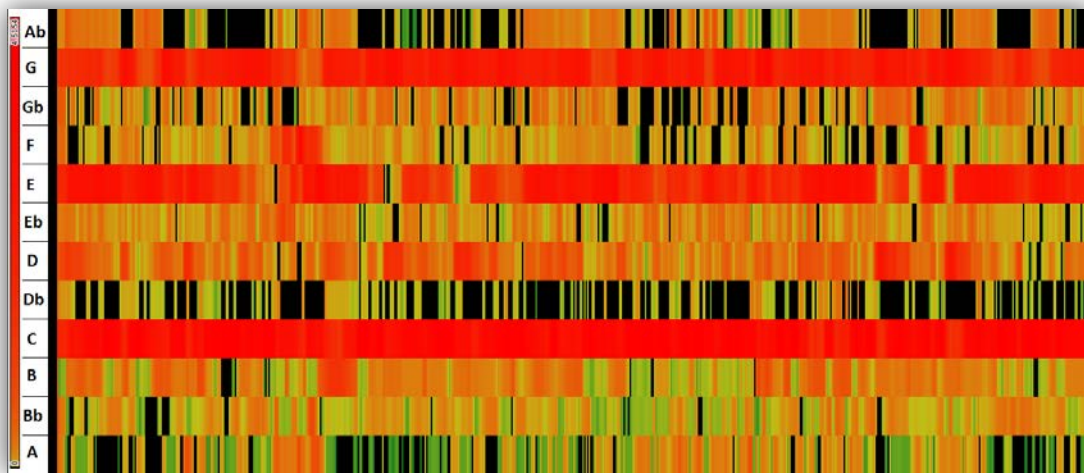
En el piano Rhodes se han tocado las tres notas que forman el acorde de manera simultánea. C, E y G (Do, Mi y Sol).



*Figura 27. Cmaj tocado en una guitarra.*

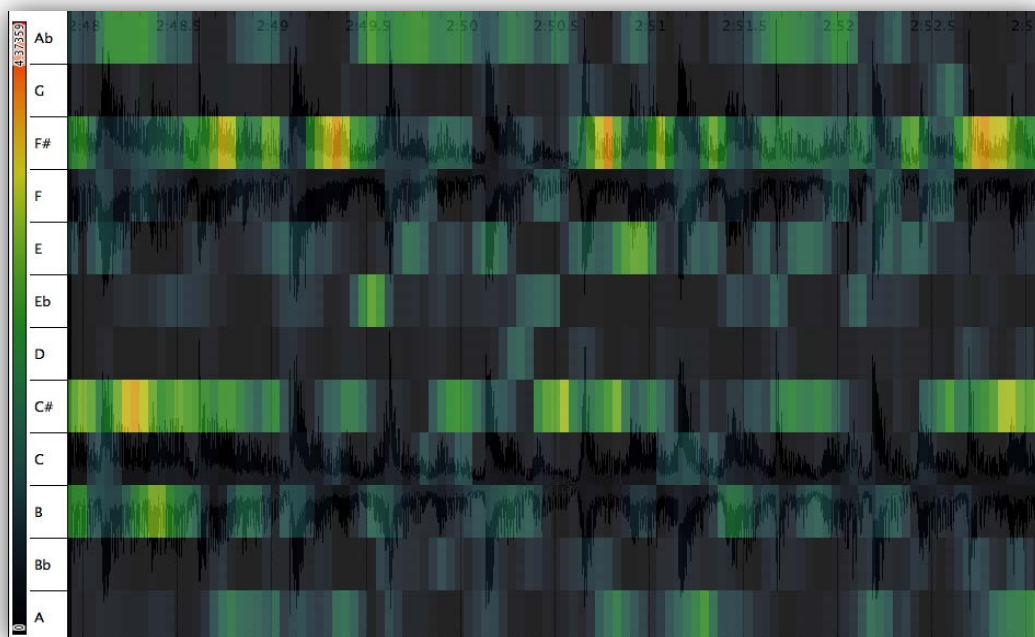
En la guitarra se han tocado las tres notas que forman el acorde de manera simultánea.

El cromograma es bastante similar al del piano Rhodes. Las ligeras diferencias frecuenciales vienen marcadas por el timbre y la forma de ejecución del acorde propia de cada instrumento.



*Figura 28. Cmaj tocado en un Piano acústico.*

En el piano, los acordes se han ejecutado de manera arpegiada. Y, además de las tres notas principales, se tocan determinadas notas añadidas susceptibles de acompañarlos cuando se utilizan en canciones. (En Cmaj se han tocado D(Re) y el F(Fa)). Con el fin de simular la forma real de interpretación del acorde de un pianista. El cromograma se presenta con más variación que los dos anteriores, precisamente por las características que se han nombrado antes. Aun así, se aprecia la preponderancia de las componentes principales del acorde.



*Figura 29. "Billie Jean" De Michael Jackson*

En la Figura 29, se puede apreciar que la información de los acordes que suenan en una canción es menos evidente que la de un acorde aislado.

### 4.1.1 Detección de acordes aislados

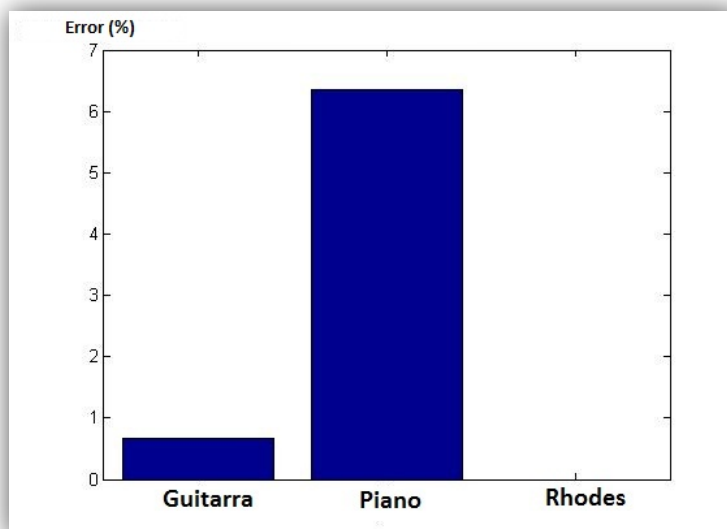
En éste primer bloque, se evaluará el reconocimiento de acordes en muestras grabadas por instrumentos. Se observará cómo se comportan ante distintos conjuntos de entrenamiento.

Los diferentes conjuntos de entrenamiento se corresponden con muestras de distintos instrumentos musicales y diversas formas de ejecución de los acordes.

Para ésta evaluación primaria se utiliza el porcentaje de error.

#### 4.1.1.1 Entrenamiento con muestras de piano Rhodes.

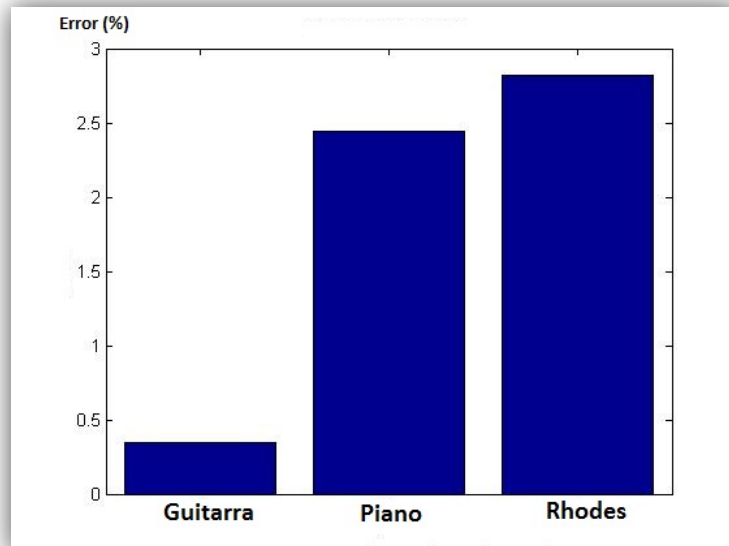
Utilizando como conjunto de entrenamiento la totalidad de las muestras grabadas con el piano Rhodes, se han obtenido estos resultados: Figura 30.



*Figura 30. Detección de acordes aislados E1.*

#### 4.1.1.2 Entrenamiento con muestras de piano acústico.

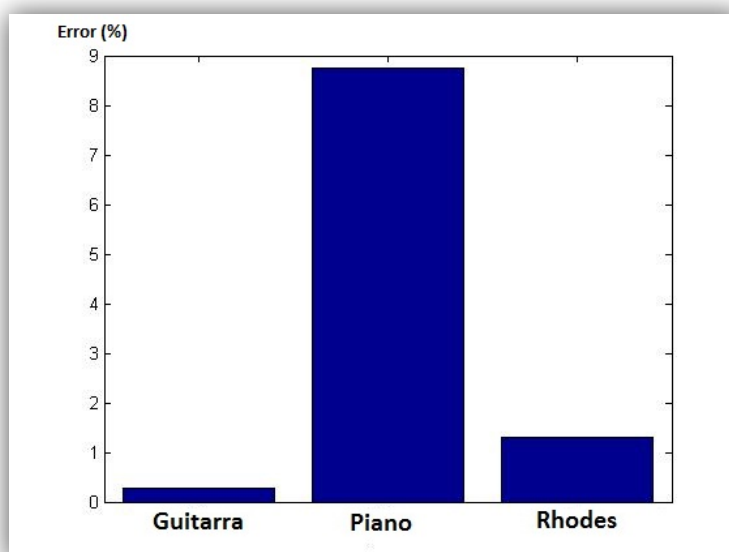
Utilizando como conjunto de entrenamiento la totalidad de las muestras grabadas con el piano acústico, se han obtenido estos resultados: Figura 31.



*Figura 31. Detección de acordes aislados E2.*

#### 4.1.1.3 Entrenamiento con muestras de guitarra.

Utilizando como conjunto de entrenamiento la totalidad de las muestras grabadas con la guitarra, se han obtenido estos resultados: Figura 32.

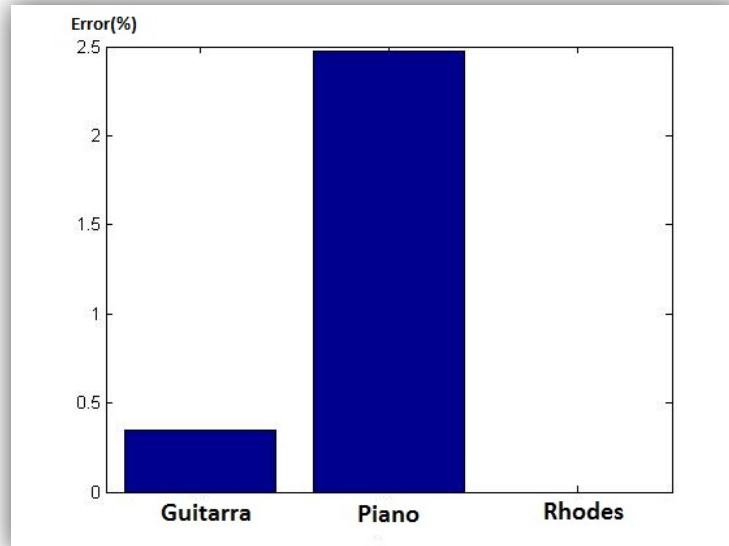


*Figura 32. Detección de acordes aislados E3.*



#### 4.1.1.4 Entrenamiento con muestras de los tres instrumentos.

Utilizando como conjunto de entrenamiento la totalidad de las muestras grabadas de todos los instrumentos, se han obtenido estos resultados: Figura 33.



*Figura 33. Detección de acordes aislados E4.*

Los resultados muestran una detección, en general satisfactoria entrenando y testeando con muestras de acordes aislados.

El conjunto de entrenamiento más eficaz, además del conjunto que engloba todas las muestras, es el de piano Rhodes, seguido de los acordes en guitarra y, por último, los arpeggios en piano acústico.

Las muestras de piano son más complicadas de reconocer, probablemente no sea por las características del instrumento, sino por la forma en la que se han tocado los acordes en el piano (apartado 4.1). Teniendo en cuenta que cada fragmento de audio a estimar, tiene una duración aproximada de 0.05 sec, hay fragmentos en los que no suenan todas las notas del acorde y puede dar lugar a estimaciones erróneas.

#### 4.1.2 Detección de acordes en canciones:

En este bloque se evaluará el reconocimiento de acordes en canciones. Como se ha hecho anteriormente con acordes aislados, se mostrarán los resultados de detección ante distintos conjuntos de entrenamiento.

La forma de presentar los experimentos será mediante un diagrama de barras. Cada barra representa el porcentaje de error en la detección de una muestra determinada.

El orden de las muestras es:

X	Muestra
1	<i>Acordes aislados</i>
	<b>Michael Jackson:</b>
2	<i>Beat it</i>
3	<i>Ben</i>
4	<i>Black or White</i>
5	<i>Blame it on the boogie</i>
6	<i>Earth Song</i>
7	<i>Heal The World</i>
8	<i>I Just can't stop loving you</i>
9	<i>Rockin Robbin</i>
10	<i>Thriller</i>
11	<i>Wanna be starting something</i>
12	<i>You are not alone</i>
	<b>Beatles:</b>
13	<i>I should have known better</i>
14	<i>And I love her</i>
15	<i>Tell me why</i>
16	<i>Any time at all</i>
17	<i>I'll cry instead</i>
18	<i>Things we said today</i>
19	<i>Ill be back</i>
20	<i>Come together</i>
21	<i>Here comes the sun</i>
22	<i>Mean mr. Mustard</i>
23	<i>Polythene pam</i>
24	<i>Twist and shout</i>
25	<i>In my life</i>
26	<i>Yellow submarine</i>
27	<i>Let it be</i>
28	<i>I need you</i>
29	<i>Help</i>

Tabla 10. Orden de las muestras en el diagrama de barras

#### 4.1.2.1 Entrenamiento con acordes aislados.

Utilizando como conjunto de entrenamiento la totalidad de las muestras grabadas de instrumentos, se han obtenido estos resultados. Figura 34.

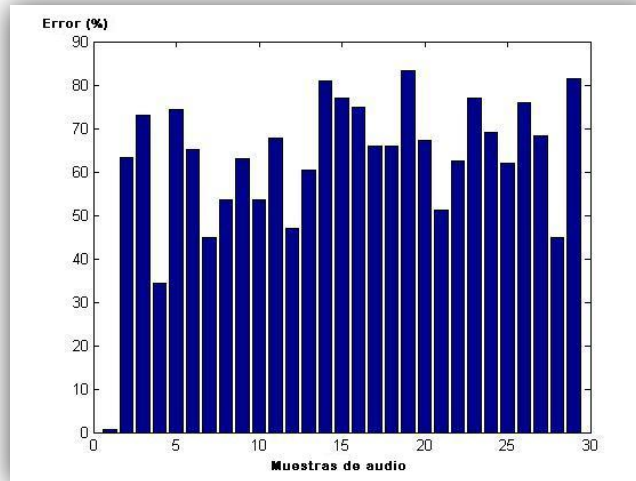


Figura 34. Detección acordes en canciones E1

#### 4.1.2.2 Entrenamiento con acordes aislados y Beatles

Utilizando como conjunto de entrenamiento la totalidad de las muestras grabadas de instrumentos y el 40% de los fragmentos de canciones de los Beatles se han obtenido estos resultados. Figura 35.

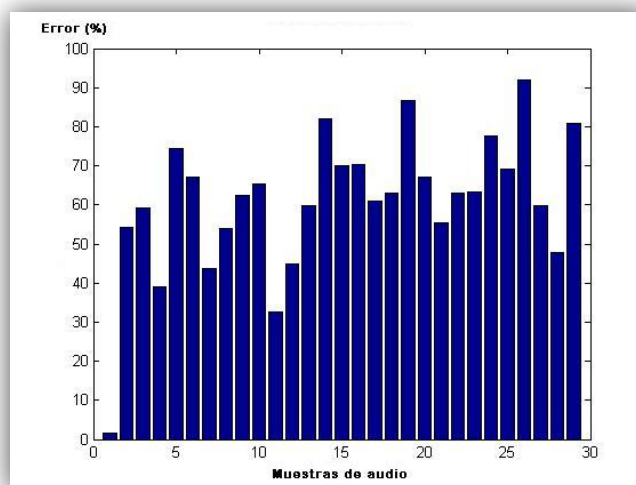


Figura 35. Detección de acordes en canciones E2

#### 4.1.2.3 Entrenamiento con acordes aislados y Michael J.

Utilizando como conjunto de entrenamiento la totalidad de las muestras grabadas de instrumentos y el 40% de los fragmentos de canciones de Michael Jackson se han obtenido estos resultados. Figura 36.

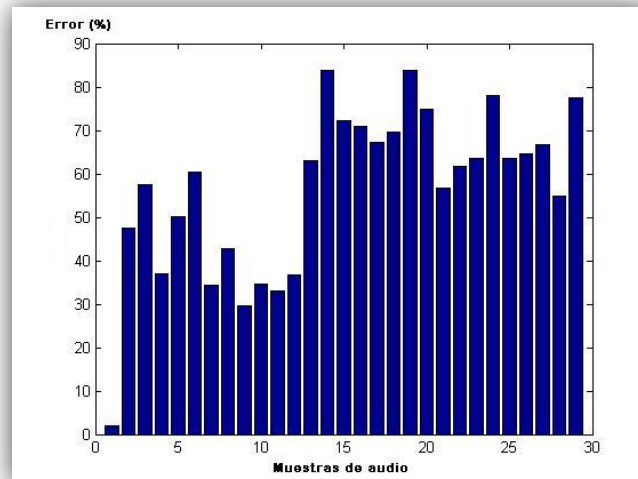


Figura 36. Detección de acordes en canciones E3.

#### 4.1.2.4 Entrenamiento con Beatles

Utilizando como conjunto de entrenamiento la totalidad de las muestras de las canciones de los Beatles se han obtenido estos resultados. Figura 37.

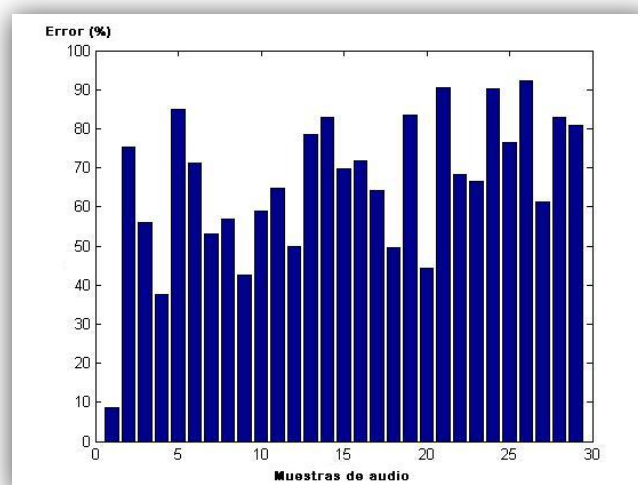


Figura 37. Detección de acordes en canciones E4

#### 4.1.2.5 Entrenamiento con Michael J.

Utilizando como conjunto de entrenamiento la totalidad de las muestras de las canciones de Michael Jackson se han obtenido estos resultados. Figura 38.

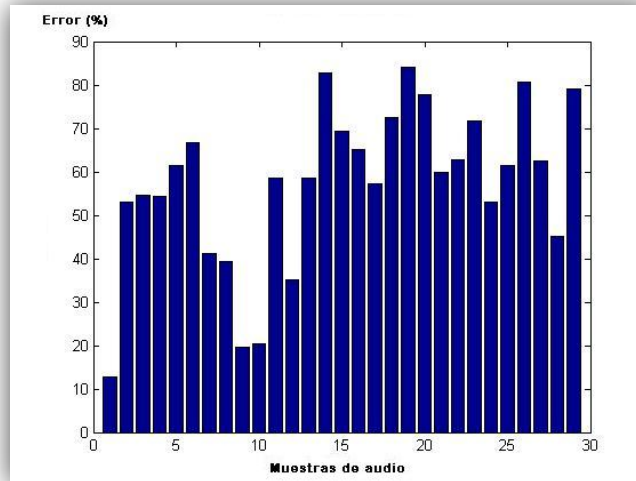


Figura 38. Detección de acordes en canciones E5

#### 4.1.2.6 Entrenamiento con Michael J. y Beatles.

Utilizando como conjunto de entrenamiento el 40% de los fragmentos de canciones de los Beatles y el 40% de los fragmentos de canciones de Michael Jackson se han obtenido estos resultados. Figura 39.

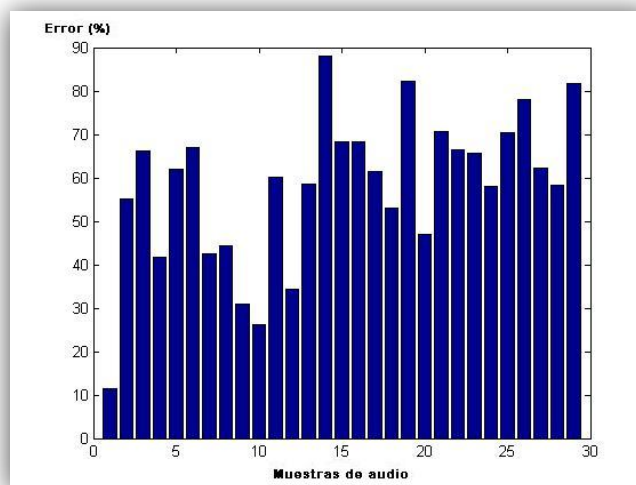


Figura 39. Detección de acordes en canciones E6

#### 4.1.2.7 Entrenamiento con acordes aislados, Michael J. y Beatles.

Utilizando como conjunto de entrenamiento la totalidad de las muestras grabadas, el 40% de los fragmentos de canciones de los Beatles y el 40% de los fragmentos de canciones de Michael Jackson se han obtenido estos resultados. Figura 40.

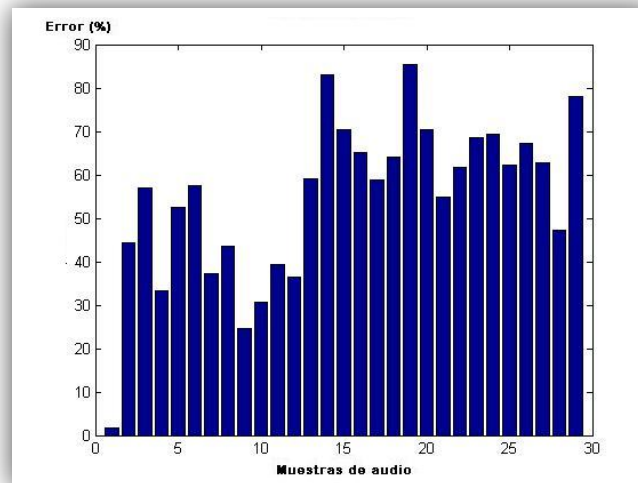


Figura 40. Detección de acordes en canciones E7.

#### 4.1.2.8 Observaciones.

En la página siguiente se muestran los resultados de los diagrama de barras anteriores en forma de tabla. Tabla 11

<b>Muestra</b>	<b>E1: Acordes</b>	<b>E2: Acordes y Beatles</b>	<b>E3: Acordes y M.J.</b>	<b>E4: Beatles</b>	<b>E5: M.J.</b>	<b>E6: Beatles y M.J.</b>	<b>E7:Beatles, M.J y Acordes</b>
<i>Acordes</i>	0.94	1.47	1.89	8.72	12.66	11.34	1.80
<i>Beat it</i>	63.19	54.18	47.56	75.32	52.89	55.15	44.26
<i>Ben</i>	72.97	59.13	57.50	55.86	54.47	66.10	57.06
<i>Black or White</i>	34.33	39.16	36.87	37.44	54.40	41.65	33.31
<i>Blame it on the boogie</i>	74.33	74.46	50.15	84.85	61.52	62.08	52.38
<i>Earth Song</i>	65.06	67.08	60.38	71.19	66.71	66.89	57.52
<i>Heal The World</i>	44.74	43.65	34.38	53.17	41.28	42.47	37.28
<i>I Just can't stop loving you</i>	53.53	54.03	42.86	56.87	39.38	44.41	43.66
<i>Rockin Robbin</i>	62.94	62.53	29.57	42.63	19.52	31.02	24.58
<i>Thriller</i>	53.63	65.44	34.63	58.86	20.43	26.16	30.61
<i>Wanna be starting something</i>	67.76	32.67	33.09	64.90	58.61	60.08	39.31
<i>You are not alone</i>	47.05	44.74	36.61	49.88	35.22	34.31	36.44
<b>Media(Michael J.)</b>	<b>58.14</b>	<b>54.28</b>	<b>42.15</b>	<b>59.18</b>	<b>45.86</b>	<b>48.21</b>	<b>41.49</b>
<i>I should have known better</i>	60.38	59.75	63.12	78.60	58.62	58.50	59.13
<i>And I love her</i>	80.99	82.14	83.88	82.82	82.78	87.92	82.90
<i>Tell me why</i>	77.06	69.89	72.13	69.59	69.24	68.34	70.33
<i>Any time at all</i>	74.87	70.21	71.04	71.67	65.01	68.41	65.01
<i>I'll cry instead</i>	65.82	60.99	67.15	64.25	57.24	61.41	58.81
<i>Things we said today</i>	65.98	63.00	69.66	49.67	72.52	52.90	64.18
<i>I'll be back</i>	83.37	86.55	83.86	83.46	84.18	82.12	85.38
<i>Come together</i>	67.10	67.03	74.87	44.33	77.80	46.84	70.33
<i>Here comes the sun</i>	51.28	55.29	56.81	90.55	59.89	70.55	54.94
<i>Mean mr. Mustard</i>	62.57	62.95	61.61	68.23	62.66	66.41	61.80
<i>Polythene pam</i>	76.98	63.22	63.49	66.66	71.82	65.60	68.51
<i>Twist and shout</i>	69.04	77.51	78.00	90.12	53.02	58.09	69.46
<i>In my life</i>	61.88	69.03	63.42	76.59	61.52	70.31	62.19
<i>Yellow submarine</i>	75.84	91.96	64.71	92.28	80.70	78.01	67.20
<i>Let it be</i>	68.21	59.69	66.76	61.17	62.38	62.20	62.88
<i>I need you</i>	44.73	47.77	54.83	82.93	45.16	58.21	47.13
<i>Help</i>	81.45	80.81	77.51	80.72	79.07	81.63	78.11
<b>Media(Beatles)</b>	<b>67.88</b>	<b>68.69</b>	<b>68.99</b>	<b>73.74</b>	<b>67.27</b>	<b>66.91</b>	<b>66.37</b>
<b>Media</b>	<b>62.34</b>	<b>60.91</b>	<b>56.49</b>	<b>65.98</b>	<b>57.27</b>	<b>57.90</b>	<b>54.71</b>

Tabla 11. Resumen de la detección

a) **Acordes aislados.**

Los resultados muestran, para todos los conjuntos de entrenamiento, que la estimación de acordes aislados es mucho más satisfactoria que la de acordes en canciones. Se puede observar, que la tasa de error más alta se produce cuando el conjunto de entrenamiento está compuesto exclusivamente con muestras de canciones. Al entrenar con canciones de los Beatles, la tasa de error es ligeramente menor que si se entrena con canciones de Michael. En el apartado 4.4 se presentan las matrices de confusión y se da una posible explicación a estos fenómenos.

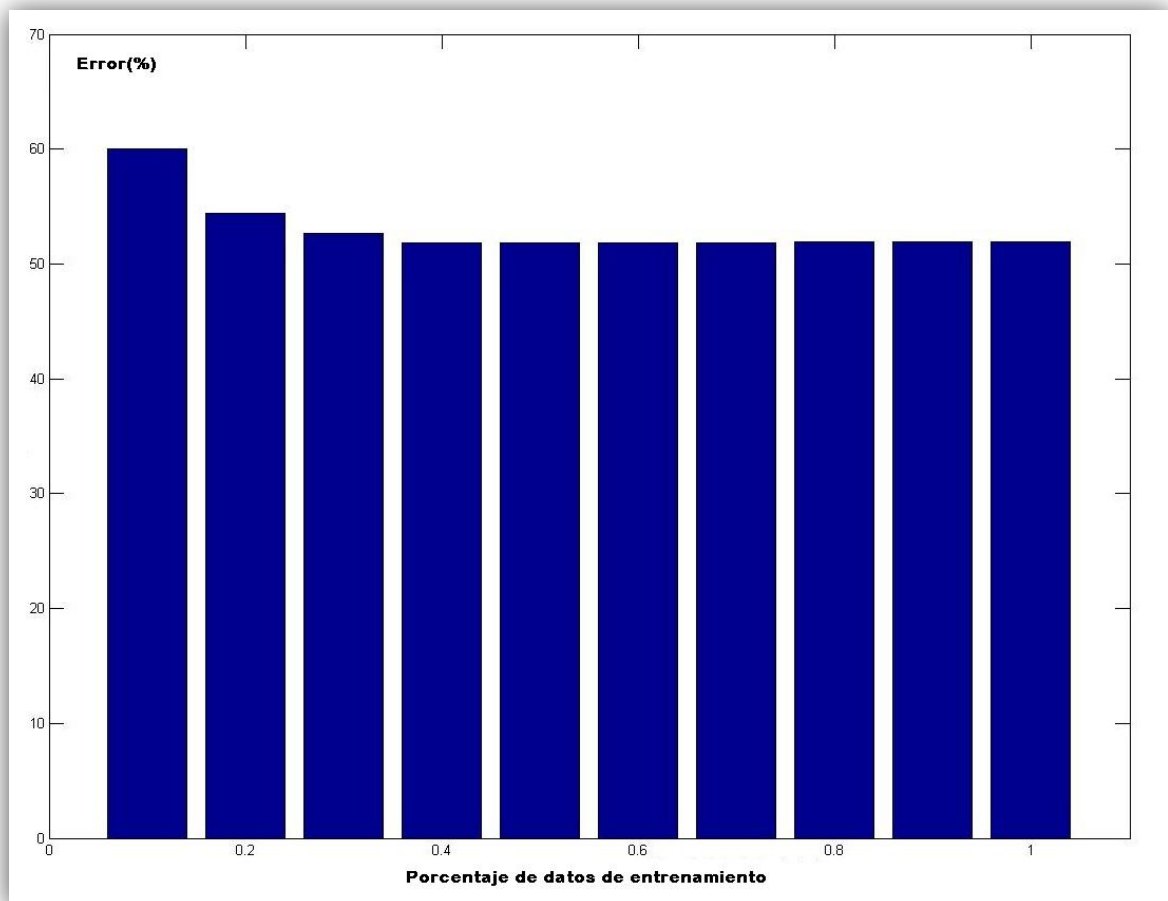
b) **Acordes en canciones.**

La tasa media de error en el caso de los temas de Michael Jackson y Beatles no varía de una forma muy destacada entre conjuntos de entrenamiento, salvo en determinados temas. La tasa media en el caso de los Beatles es generalmente mayor. En la conclusión se dan posibles causas que provoquen éste y otros comportamientos del detector.



## 4.2 Influencia del porcentaje del conjunto de entrenamiento

En este bloque se va a evaluar cómo influye, en el porcentaje de error, el porcentaje de datos de entrenamiento usados para entrenar el detector.



*Figura 41. Influencia del porcentaje de entrenamiento.*

La gráfica muestra una disminución en la tasa de error a medida que aumentamos el porcentaje de datos de entrenamiento. Sin embargo, a partir del 40% el porcentaje de error se mantiene estable. El valor óptimo es 0.4.

### 4.3 Influencia del postprocesado:

El postprocesado, como se ha indicado en el capítulo anterior, suaviza las transiciones que existen entre acordes en cortos periodos de tiempo.

La primera prueba va a consistir en conocer la influencia que tiene éste en el porcentaje de error de clasificación

Cada diagrama de barras representa, para un determinado conjunto de entrenamiento el porcentaje de error en la detección de acordes de diversas muestras de audio. Se han hecho dos tipos de detecciones para cada muestra. En una de ellas se ha prescindido del bloque de suavizado y en la otra se ha utilizado.

En la Figura 1 se ha realizado la detección de acordes en muestras de instrumentos. Guitarra, piano acústico y piano Rhodes.

El utilizar el bloque de postprocesado la tasa de error ha disminuido en promedio más de un 50% en el caso de acordes aislados. Figura 42.

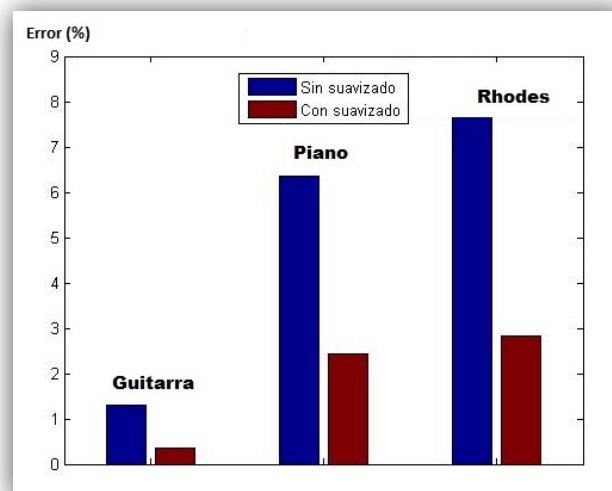


Figura 42. Postprocesado E1.

Las Figuras 43 y 44 representan la detección en un conjunto de canciones utilizando dos conjuntos de entrenamiento diferentes.

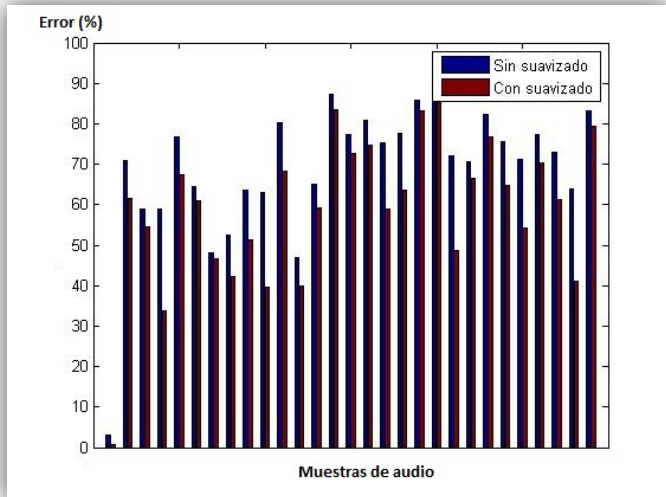


Figura 43. Postprocesado E2

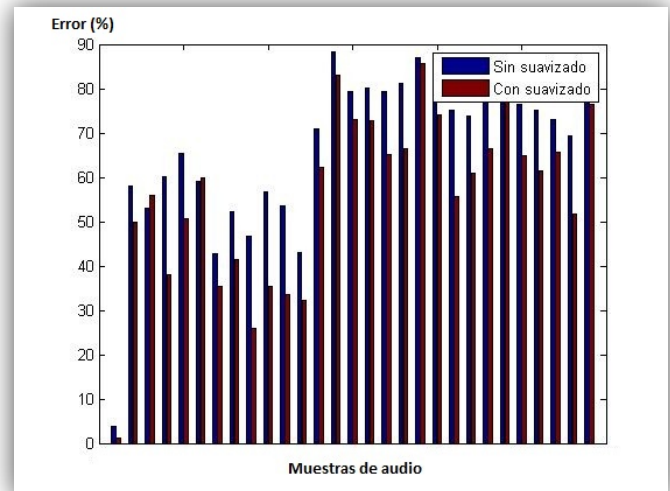


Figura 44: Postprocesado E3.

En el caso de canciones, la tasa de error disminuye un promedio del 10% al utilizar el bloque de postprocesado.

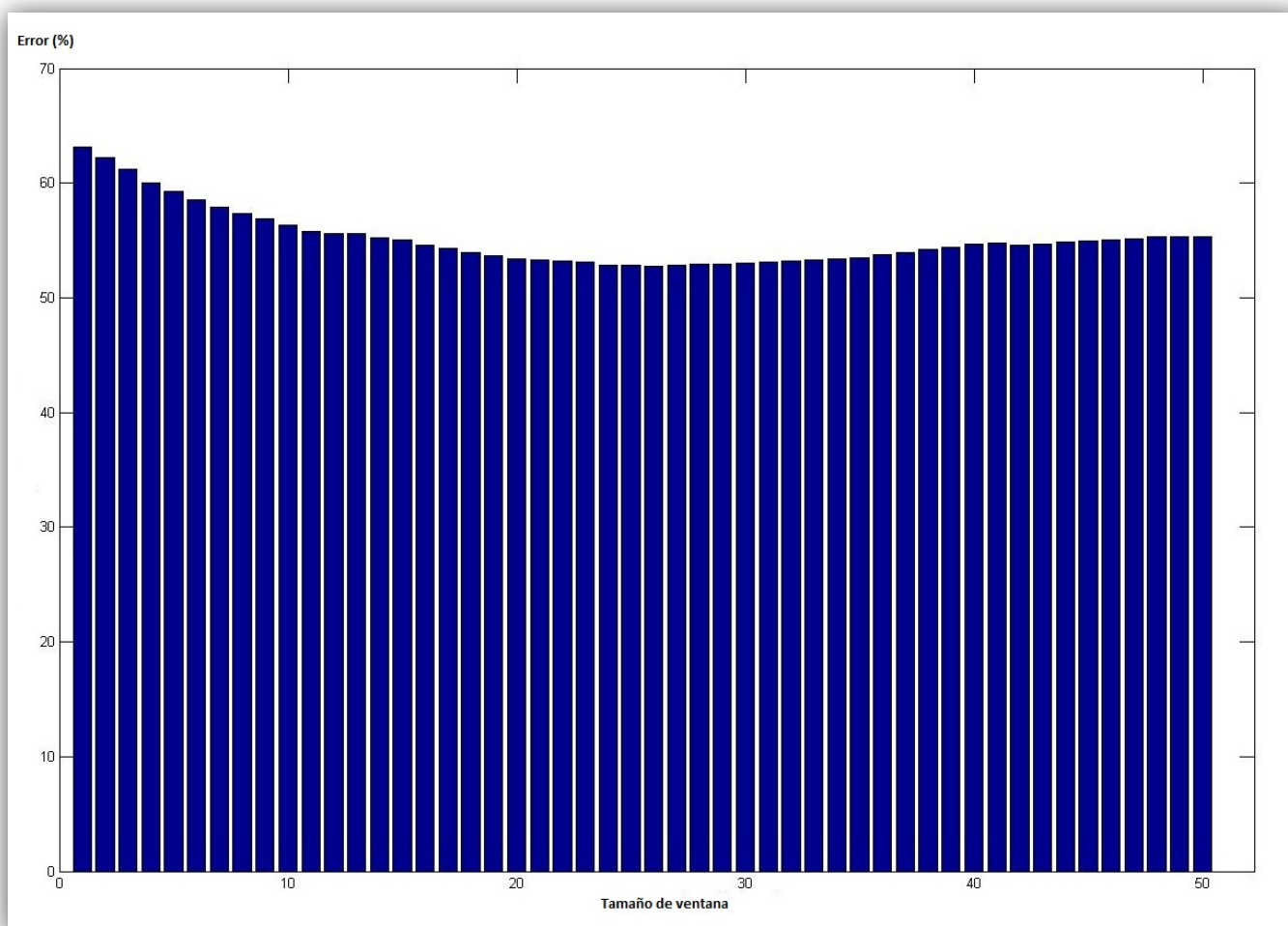
Cabe mencionar que, en los experimentos de los apartados anteriores ya se había incluido este bloque.

#### 4.3.1 Tamaño de la ventana.

Se ha mencionado en el apartado 3.3.3.2 que el bloque de suavizado calcula la moda sobre un determinado conjunto de muestras. El número de muestras de ese conjunto viene determinado por el tamaño de ventana.

El siguiente paso tiene como objetivo hallar el tamaño de ventana óptimo que minimiza el porcentaje el error.

Sobre un conjunto de entrenamiento determinado se ha calculado el porcentaje de error medio con distintos tamaños de ventana. Figura 45



*Figura 45. Influencia del tamaño de ventana del postprocesado.*

Se puede observar que el error disminuye con un tamaño de ventana entre 23 y 30 muestras aproximadamente.

También cabe destacar que el valle que se produce en el rango central de la función es poco abrupto. Esta característica indica la poca sensibilidad del porcentaje de error ante cambios sustanciales en el tamaño de ventana dentro de ese rango.

Para los experimentos siguientes, se asume la utilización del postprocesado con un tamaño de ventana de 26 muestras, uno de los valores con los que se minimiza el error.

## 4.4 Matrices de confusión

En este último bloque se evalúa la detección utilizando la información que proporciona la matriz de confusión. La forma de representarlas es mediante el *diagrama de Hinton*. Este método consiste en presentar de forma gráfica los datos numéricos de una matriz, en la cual, cada elemento está asociado a un cuadrado cuya área es proporcional a la magnitud de éste.

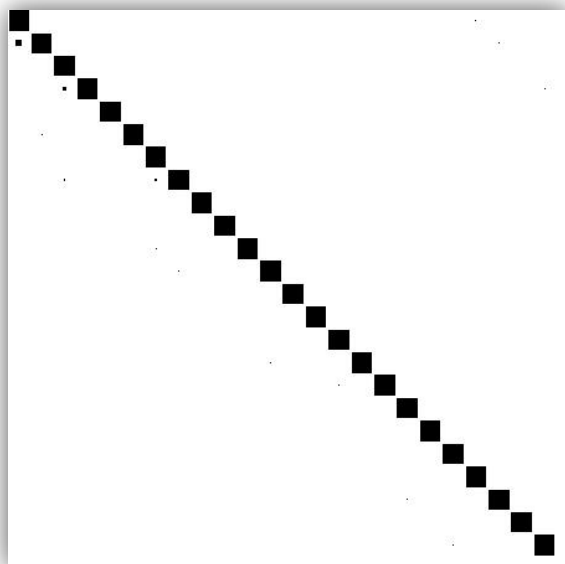
El orden tanto de las filas y columnas es:

Nº fila/columna	Acorde
1	Amaj
2	Amin
3	Bbmaj
4	Bbmin
5	Bmaj
6	Bmin
7	Cmaj
8	Cmin
9	Dbmaj
10	Dbmin
11	Dmaj
12	Dmin
13	Ebmaj
14	Ebmin
15	Emaj
16	Emin
17	Fmaj
18	Fmin
19	Gbmaj
20	Gbmin
21	Gmaj
22	Gmin
23	Abmaj
24	Abmin
25	No acorde

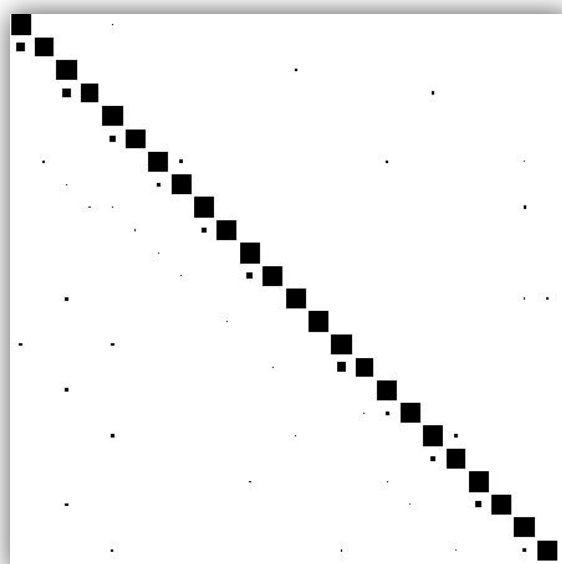
Tabla 12. Orden de filas y columnas de la Matriz de confusión.

#### 4.4.1 En acordes aislados

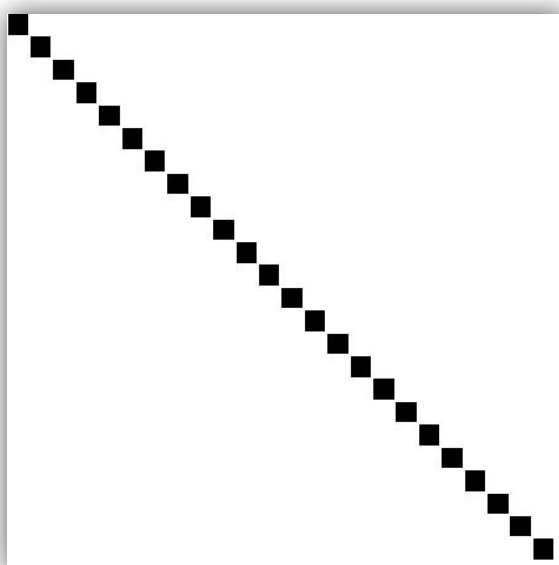
##### 4.4.1.1 Entrenamiento con muestras de piano Rhodes:



*Figura 46. M.Conf de muestras de guitarra.*

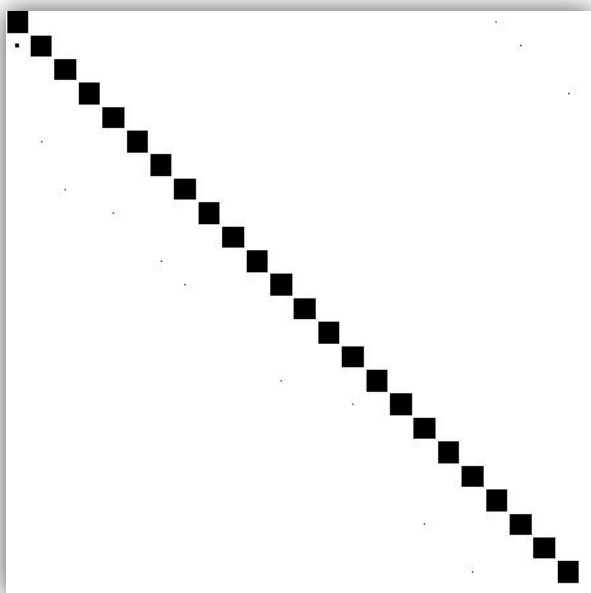


*Figura 47. M.Conf de muestras de piano.*

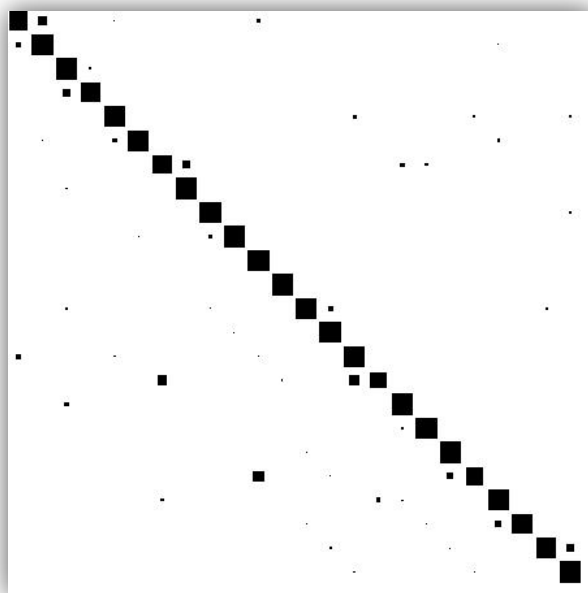


*Figura 48. M.Conf de muestras de Rhodes*

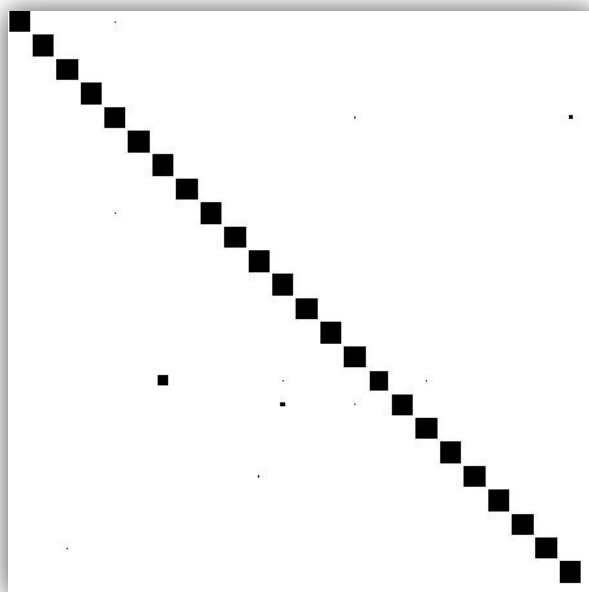
#### 4.4.1.2 Entrenamiento con muestras de guitarra.



*Figura 49. M.Conf de muestras de guitarra.*

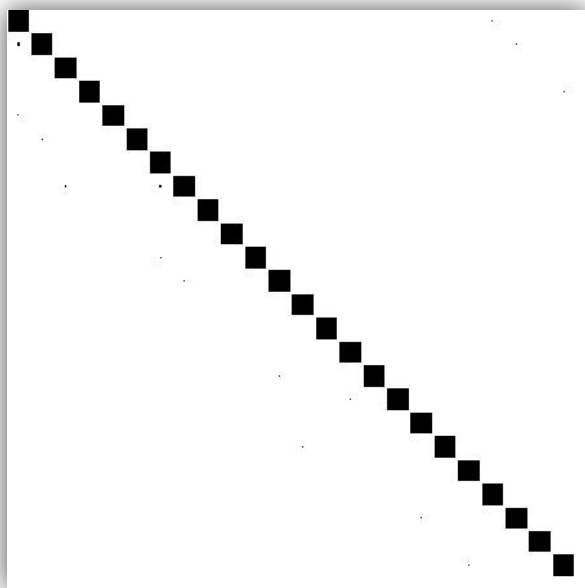


*Figura 50. M.Conf de muestras de piano.*

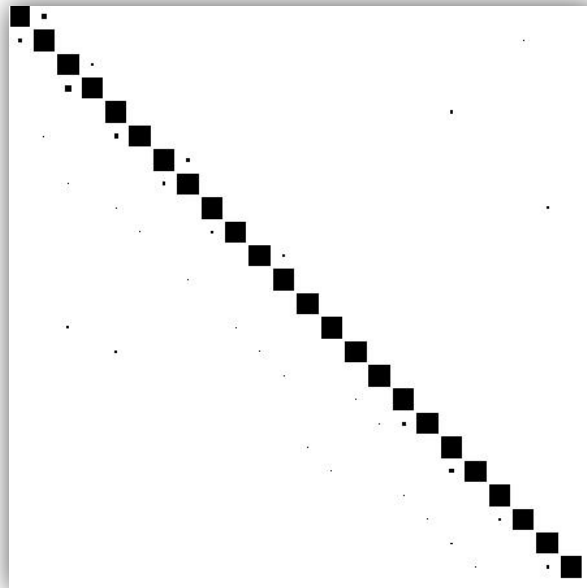


*Figura 51. M.Conf de muestras de piano Rhodes.*

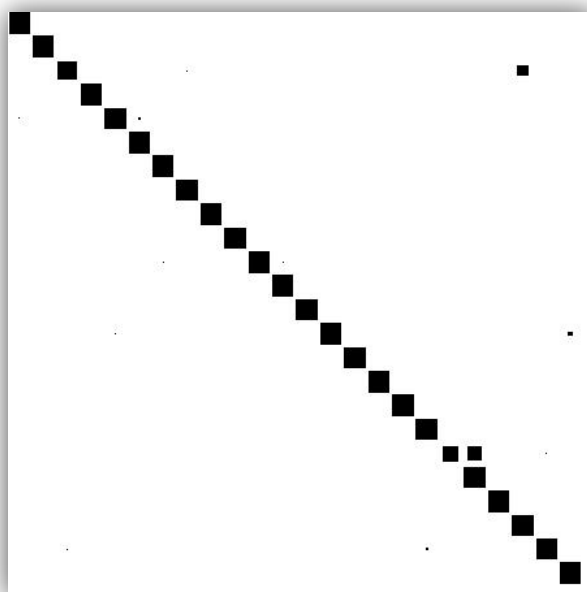
#### 4.4.1.3 Entrenamiento con muestras de piano acústico.



*Figura 52. M.Conf de muestras de guitarra*



*Figura 53. M.Conf de muestras de piano.*



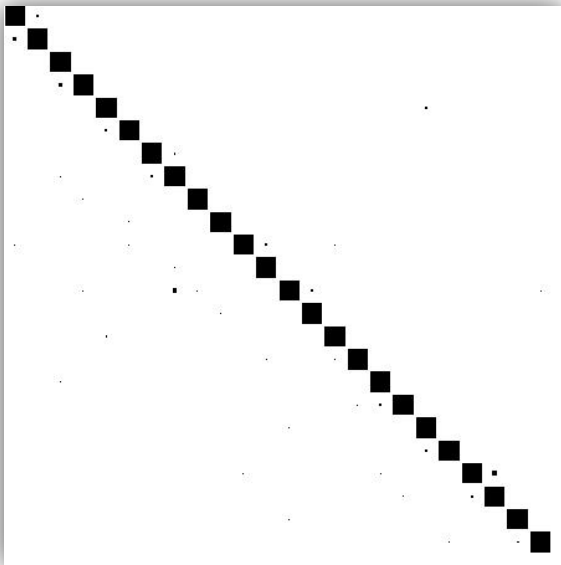
*Figura 54. M.Conf de muestras de Rhodes*



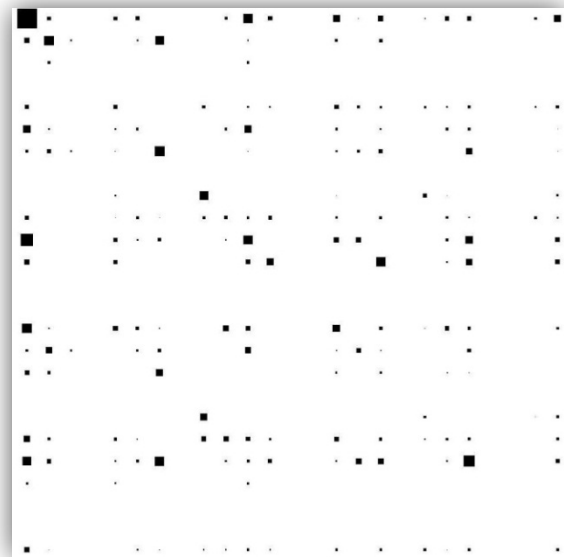
Las observaciones muestran la confusión, en algunos casos, del modo mayor y menor de algunos acordes bajo una misma tónica. (Véase apartado 2.1.2 sobre la descripción de los modos de acorde). Esto es producido por la similitud que existe entre éstos, ya que comparten dos de sus tres notas, y entre las notas que difieren, existe una distancia de un semitono, es decir, están muy próximas en términos de frecuencia. Este fenómeno se produce algo más en el caso del piano, debido a que, al tocarse las notas de manera sucesiva y no simultánea, en determinados fragmentos sólo suenan las dos notas comunes entre estos dos acordes, y puede inducir a error en la detección.

#### 4.4.2 Acordes en canciones

##### 4.4.2.1 Entrenamiento con acordes aislados, Michael J. y Beatles.



*Figura 55. M.Conf de acordes aislados*



*Figura 56. M.Conf de Beatles*

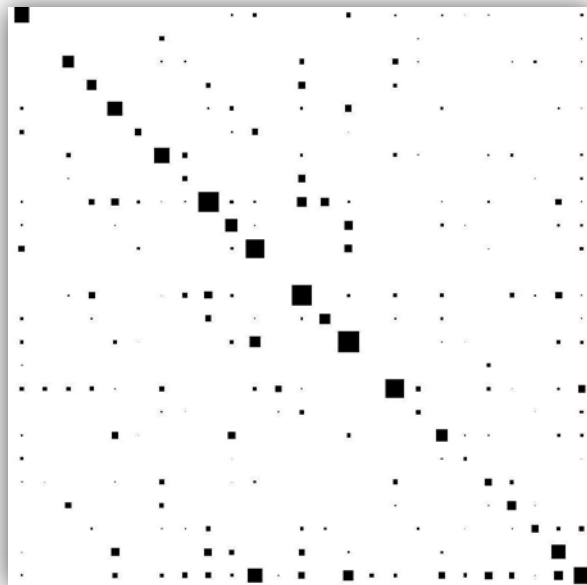


Figura 57. M.Conf de muestras de Michael Jackson

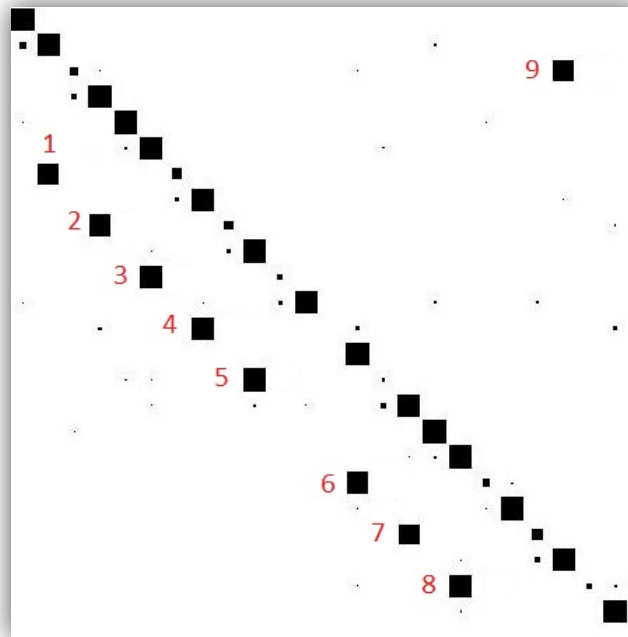
Se puede observar, que tanto en Michael Jackson como con los Beatles, hay acordes que son más frecuentes que otros. De hecho, hay algunos que no aparecen, ya sea porque su utilización es menos frecuente o por mera casualidad en las canciones escogidas. En cualquier caso, se puede observar que el reconocimiento ha dado mejores resultados en los temas de Michael Jackson (Figura 57) que en los temas de los Beatles (Figura 56). En éstos no se ha conseguido un reconocimiento del todo fiable.

A continuación se muestran dos matrices de confusión que pueden resultar interesantes. El reconocimiento de los acordes aislados con el entrenamiento únicamente de las muestras de audio de las canciones. Cabe mencionar, que hay determinados acordes que no aparecen en las muestras de canciones. En el caso de Michael Jackson, sólo se da en un caso: *Dmin*.

En las canciones de los Beatles, a pesar de que se dispone de más canciones, se produce en muchas más clases: *Bbmin*, *Cmin*, *Ebmaj*, *Ebmin*, *Fmin*, *Gbmaj*, *Abmaj* y *Abmin*.

Para tales acordes, se han utilizado las muestras grabadas para inicializarlos, hecho que puede influir en el reconocimiento de esas clases.

#### 4.4.2.2 Entrenamiento con muestras de Michael J.



*Figura 58. M.Conf de muestras de acordes aislados*

1. Cmaj,Amin
2. Dbmaj,Bbmin
3. Dmaj,Bmin
4. Ebmaj,Cmin
5. Emaj,Dbmin
6. Gbmaj,Ebmin
7. Gmaj,Emin
8. Abmaj,Fmin
9. Bbmaj,Gmin

La Figura 58 muestra el reconocimiento erróneo de las clases enumeradas en la parte superior. Esta confusión no es aleatoria, ya que se da sólo en acordes mayores. Además se observa en estos casos que cada acorde detectado está a una misma distancia relativa del acorde etiquetado. Esta confusión se produce siempre con el acorde menor que se encuentra a 3 semitonos en sentido descendente del acorde mayor. Si tenemos en cuenta explicaciones previas (apartado 2.1.4) se concluye que los acordes estimados erróneamente son relativos del acorde asociado a la observación.

Pasa lo mismo en el caso número 8 de la misma figura. En éste se confunde el acorde *Bbmaj* con *Gmin*. Aunque visualmente se encuentre a una distancia mucho mayor, debido a la propiedad de periodicidad de las notas (apartado 2.1.1) la distancia entre tónicas también es de 3 semitonos.

#### 4.4.2.3 Entrenamiento con muestras de Beatles

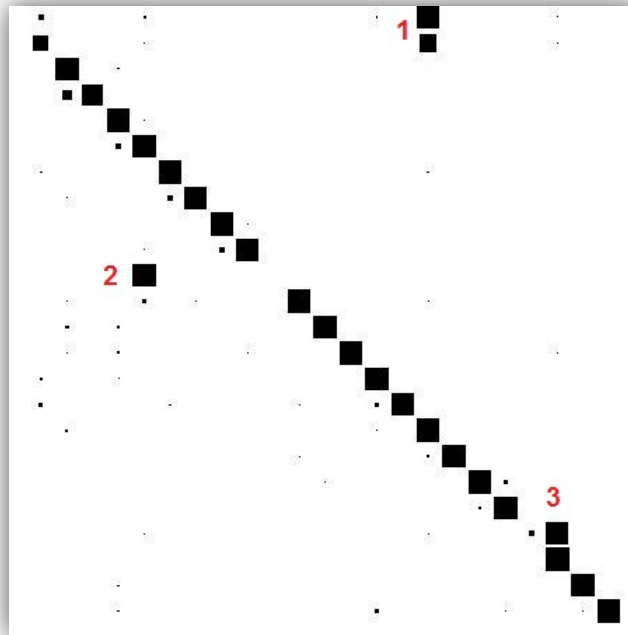


Figura 59. M.Conf de muestras de acordes aislados

1. Confusión entre Amaj y Fmin y entre Amin y Fmin. Como se ha detallado al inicio de este apartado, la clase que corresponde al acorde de Fmin ha sido entrenada con muestras de acordes aislados ya que no aparece en las canciones de los Beatles usadas. Puede ser que este hecho provoque tal confusión.
2. Confusión entre Dmaj y Bmin. Tal y como se explica en el ejemplo anterior, los dos acordes confundidos son relativos entre sí.
3. Confusión entre Gmaj y Gmin. Al igual que pasaba en las muestras de acordes aislados, aunque en mayor medida, se produce una confusión entre el modo mayor y menor.

## 4.5 Valoración de las pruebas.

Tras la evaluación de los resultados se puede afirmar que se produce una detección satisfactoria en el caso de los acordes aislados (apartado 4.1.1).

En el caso de canciones, es necesario un desarrollo más sofisticado para conseguir resultados fiables y contrastar las conclusiones obtenidas.

Hay que tener en cuenta que las muestras etiquetadas, para el entrenamiento de un conjunto de 25 clases, son escasas y poco homogéneas (En las canciones hay acordes que aparecen mucho más que otros). Este hecho puede producir una variación en el porcentaje de error de detección de determinadas canciones al escoger conjuntos de entrenamiento insuficientes (apartado 4.1.2.8)

Aun así, se reconocen patrones característicos en la detección, como la confusión entre acordes relativos y modos mayores y menores (Figura 58 y Figura 59). También se demuestra la influencia del postprocesado en forma de filtro de modas para conseguir un resultado más fiable (apartado 4.3).

Por último, la influencia también satisfactoria que ha tenido utilizar muestras de acordes aislados en el conjunto de entrenamiento, ya sea para el reconocimiento de éstos, o el reconocimiento de acordes en canciones. (apartado 4.1.2.8)



## **Capítulo 5. Conclusión**

Tras el desarrollo del trabajo descrito en los capítulos precedentes, se exponen las conclusiones e ideas a destacar.

En primer lugar, se amplía la base de datos existente en la página web [isophonics.net] con el etiquetado de canciones de Michael Jackson y con la grabación y etiquetado de muestras de audio de instrumentos aislados para posteriormente evaluar su contribución.

Tras la implementación del detector, se evalúa el reconocimiento de acordes en muestras de acordes aislados ejecutados en piano acústico, piano Rhodes y guitarra. La tasa de error varía en función del conjunto de entrenamiento utilizado, pero en promedio se encuentra por debajo del 5%. Se puede considerar una detección satisfactoria.

En los resultados de la evaluación del reconocimiento de acordes en canciones, el porcentaje de error promedio se ha conseguido reducir a un 41% en el caso de Michael Jackson y un 66% en el caso de los Beatles. Teniendo en cuenta que el porcentaje de error aleatorio es del 96%, se puede asumir que se ha producido una contribución positiva al problema del reconocimiento, aunque es necesaria una mejora sustancial para obtener resultados fiables.

Se corrobora sin embargo, que la utilización de acordes aislados en el conjunto de entrenamiento, mejora la detección de estos cuando se encuentran en canciones.

Los resultados también muestran una contribución positiva del bloque de postprocesado. Se ha logrado reducir el porcentaje de error más de un 50% en el caso del reconocimiento de acordes aislados y alrededor del 10% en el caso de acordes en canciones. Además los resultados determinan que, para minimizar el error, el tamaño de ventana sobre el que se calcula el postprocesado debe estar entre 23 y 30 fragmentos detectados.

Por último, mediante el uso de la matriz de confusión, se logran identificar patrones que se producen en la estimación errónea de acordes en canciones. Tales como la confusión entre el modo mayor y menor o la confusión entre acordes relativos.

## 5.1 Trabajo futuro.

El desarrollo del prototipo por bloques separados permite la mejora independiente de cada uno de ellos con el fin de realizar un reconocimiento más eficaz.

A continuación se propondrán mejoras en base a las dificultades y necesidades encontradas a lo largo de la implementación del prototipo.

En lo que concierne al primer bloque, la base de datos es reducida y poco heterogénea para un número de clases tan alto. Hay acordes que, en el conjunto de las canciones escogidas, aparecen en menor medida, y provoca que no haya suficientes muestras para que se produzca un entrenamiento deseado.

Es necesario por lo tanto, el **aumento de la base de datos** con canciones etiquetadas de otros grupos musicales, estilos, etc. Y sobre todo, se deben incluir muestras de acordes aislados tocados por distintos instrumentos que contribuyen positivamente en el reconocimiento.

En la página web consultada [isphonics.net] no existen como tal, y sólo se puede acceder a una base de datos de acordes en canciones.

Otra dificultad encontrada es la de la reducción de acordes de más de tres notas a acordes de tres notas, que son los utilizados en el proyecto. A pesar de que, teóricamente, un acorde de cuatro o cinco notas puede reducirse a uno de tres (apartado 3.2.3), se necesita una clasificación separada para realizar una transcripción más exhaustiva.

Por lo tanto, otra posible mejora sería la **ampliación del número de clases** para incluir acordes de 4 notas.

En el bloque de detección, se pueden desarrollar algoritmos de detección más sofisticados. Como se ha indicado en el apartado 2.2.2.2, en los últimos trabajos se han implementado algoritmos asociados a los **modelos de Markov** para tener en cuenta la probabilidad de transición entre estados.

En teoría musical, la transición entre acordes se denomina cadencia. Hay cadencias más utilizadas que otras, y por lo tanto más probables [Herrera 1989]. La



consideración de las cadencias musicales, y su desarrollo en modelos de Markov pueden contribuir a la mejora en la detección.

Además de las mejoras propuestas para lograr una detección más fiable, puede continuarse el proyecto por diferentes vías:

- **Desarrollo de un interfaz gráfico** que permita al usuario introducir una muestra de audio y conocer el acorde que se ejecuta en cada momento de la reproducción.
- **Combinación con otro tipo de detecciones sobre un archivo de audio.** Por ejemplo, la detección de tempo y/o línea melódica se puede combinar con la detección de acordes para desarrollar una transcripción más completa de una canción.
- **Optimización del detector** para un instrumento concreto, como por ejemplo el piano o la guitarra, para el desarrollo de una herramienta pedagógica que pueda ser usada en el aprendizaje de acordes del instrumento deseado.



## **Referencias**

### **Libros:**

[Recuero 1999] *Ingeniería acústica*. Manuel Recuero López. Paraninfo, 1999.

[Tipler 2007] *Physics for Scientists and Engineers*. Paul A. Tipler and Gene Mosca. Enlarged 6<sup>th</sup> edition. 2007

[Oppenheim 1997] *Señales y Sistemas*. Alan v. Oppenheim, Alan s. Willsky, S.Hamid Nawab. Segunda edición. Prentice Hall. 1997

[Herrera 1984] *Teoría musical y Armonía Moderna Vol. I*. Enric Herrera. Editorial Antoni Bosch. 1984

[Sala 1997] *ABC del MIDI*. Pablo Sala. Editorial Ricordi. 1997

[Johansson 1998] *Real rock book*. K G Johansson. Editorial Warner. 1998.

### **Artículos y tesis:**

[Ulrich 1977] *The analysis and synthesis of jazz by computer*. John Wade Ulrich. Computing and information Science Department. University of New Mexico. 1977

[Temperley and Sleator 1999] *Modeling Meter and Harmony: A Preference-Rule Approach*. David Temperley and Daniel Sleator. *Computer music Journal*. 1999

[Fujishima 1999] *Realtime chord recognition on musical sound: A system using Common Lisp music*. Takuya Fujishima. En *Proceedings of the International Computer Music Conference(ICMC)*, páginas 464-467, 1999.

[Wakefield 1999] *Mathematical representation of joint time-chroma distributions*. Gregory H. Wakefield. En *Proceedings of SPIE*, volume 3807, 1999.

[Pardo 2002] *Algorithms for chordal analysis*. Bryan Pardo. *Computer Music Journal*. 2002

[Raphael and Stoddard 2003] *Harmonic analysis with probabilistic graphical models*. C. Raphael and J. Stoddard. En *Proceedings of the 4<sup>th</sup> International Conference on Music Information Retrieval. ISMIR 2003, Baltimore, USA*. 2003.

[Sheh and Ellis 2003] *Chord segmentation and recognition using EM-trained hidden Markov models*. Alexander Sheh and Daniel Ellis. En *Proceedings of the 4<sup>th</sup> International Conference on Music Information Retrieval. ISMIR 2003, Baltimore, USA*. 2003.

[Paiement 2005] *A probabilistic model for chord progressions*. Jean-François Paiement. En *Proceedings of the 6<sup>th</sup> International Conference on Music Information Retrieval. ISMIR 2005, London, UK*, pages 312-319, 2005.

[Harte and Sander 2005] *Symbolic representation of musical chords: A proposed syntax for text annotations*. Christopher Harte, Mark Sandler. En *Proceedings of the 6<sup>th</sup> International Conference on Music Information Retrieval. ISMIR 2005, London, UK*, pages 66-71, 2005.

[Bello and Pickens 2005] *A Robust Mid-Level Representation for Harmonic Content in Music Signals*. Juan P. Bello and Jeremy Pickens. En *Proceedings of the 6<sup>th</sup> International Conference on Music Information Retrieval. ISMIR 2005, London, UK*, pages 304-311, 2005.

[Catteau et al. 2007] *A probabilistic framework for audio-based tonal key and chord recognition*. Benoit Catteau, Jean-Pierre Martens, and Marc Leman. En *Proceedings of the 30<sup>th</sup> Annual Conference of the Gesellschaft für Klassifikation e.V., Freie Universität Berlin. ISMIR March 8-10*, pages 637-644, 2007.

[Papadopoulos and Peeters 2007] *Large-scale Study of Chord Estimation Algorithms Based on Chroma Representation and HMM*. H  lene Papadopoulos and Geoffroy Peeters. En *International Workshop on Content-Based Multimedia Indexing*, p  ginas 53-60, 2007.

[Mauch 2010] *Automatic Chord Transcription from Audio Using Computational Models of Musical Context*. Matthias Mauch. Tesis doctoral. *School of Electronic Engineering and Computer Science Queen Mary, University of London*. March 23, 2010.

**Páginas web:**

[Wikipedia.org] <http://es.wikipedia.org/wiki/Acorde>. (20, Mayo, 2012)

[isophonics.net] [http:// isophonics.net](http://isophonics.net) (12 Abril 2012)

[isophonics Beatles] <http://isophonics.net/content/reference-annotations-beatles> (12 Abril 2012)



## Presupuesto:

### Material:

- Ingeniero:**

<i>Ordenador Mac book pro</i>	<b>1.150 €</b>
<i>Licencia Matlab para estudiantes</i>	<b>99 €</b>
<i>Sonic Visualizer.</i> <i>(Programa de extracción del cromograma)</i>	<b>0€</b>
<i>Licencia Microsoft Office 2010</i>	<b>120 €</b>
<i>Canciones de Michael Jackson</i> <i>(11 canciones)</i>	<b>11 x 1€=11 €</b>
<b>TOTAL</b>	<b>1380 €</b>

- Músico:**

<i>Ordenador Mac book pro</i>	<b>1.300 €</b>
<i>Licencia Logic pro 9.</i> <i>(Secuenciador utilizado en el etiquetado)</i>	<b>150 €</b>
<i>Licencia Ivory I.</i> <i>(Samplers de piano acústico)</i>	<b>129 €</b>
<i>Licencia Fender Rhodes</i> <i>(Samplers de piano Rhodes)</i>	<b>50 €</b>
<i>Licencia Guitar Rig.</i> <i>(Emulación de amplificadores de guitarra)</i>	<b>30€</b>
<i>Teclado MAudio Oxygen.</i> <i>(Teclado que envía datos MIDI)</i>	<b>185€</b>
<i>Guitarra Fender Telecaster</i>	<b>600 €</b>
<i>Tarjeta de sonido Motu Ultralite y cableado</i>	<b>600 €</b>

El coste de material del Músico es meramente informativo y está incluido como amortización en el coste por hora de los servicios de éste. Se ha estimado en 15€/hora.

## Personal:

- **Ingeniero:**

<i>Consulta de la bibliografía</i>	<b>20h</b>
<i>Diseño del prototipo</i>	<b>20h</b>
<i>Programación del prototipo y pruebas</i>	<b>200h</b>
<i>Memoria escrita</i>	<b>100h</b>
Coste	15€/hora
<b>TOTAL</b>	<b><del>15€</del>340h= 5100 €</b>

- **Músico:**

<i>Obtención de los acordes de las canciones de Michael Jackson y etiquetado</i>	<b>10h</b>
<i>Grabación de muestras de acordes aislados y etiquetado</i>	<b>5h</b>
Coste	30 €/hora = 15€ (material)+15€(personal)
<b>TOTAL</b>	<b><del>30€</del>15h=450 €</b>

## Presupuesto total.

Material:	<b>1380 €</b>
Ingeniero:	<b>5100 €</b>
Músico:	<b>450 €</b>
<b>TOTAL</b>	<b>6930 €</b>